

СТРУКТУРА И ЭВОЛЮЦИЯ ДНК-ТРАНСПОЗОНОВ НАДСЕМЕЙСТВА
L31 ДВУСТВОРЧАТЫХ МОЛЛЮСКОВ© 2024 г. М. В. Пузаков^а, *, Л. В. Пузакова^а^а Федеральный исследовательский центр “Институт биологии южных морей им. А.О. Ковалевского” Российской академии наук, Севастополь, 299011 Россия

*e-mail: puzakov@ngs.ru

Поступила в редакцию 07.04.2023 г.

После доработки 09.06.2023 г.

Принята к публикации 09.06.2023 г.

Мобильные генетические элементы *IS630/Tc1/mariner* (*ITm*) являются широко распространенными представителями ДНК-транспозонов, вносящих значительный вклад в эволюцию геномов эукариот. Масштабное применение технологий секвенирования нового поколения (NGS) привело к появлению множества полногеномных нуклеотидных последовательностей новых организмов и выявлению элементов *ITm* в большинстве таксонов эукариот. Несмотря на достаточно детальную изученность разнообразия *ITm*, по-прежнему обнаруживаются элементы, которые способствуют расширению и пересмотру классификации этой группы ДНК-транспозонов. В представленной работе впервые проведен детальный анализ элементов *L31* двустворчатых моллюсков, что позволило описать структуру, разнообразие, распространение и филогенетическое положение этих элементов в группе *ITm*. Установлено, что *L31*-транспозоны являются самостоятельным надсемейством в группе *ITm*, имеющим древнее происхождение. Внутри клады *L31* наблюдается достаточно высокое разнообразие — выделено пять филогенетических кластеров. На данный момент *L31*-транспозоны у двустворчатых моллюсков выявлены только в подклассе Autobranchia с преобладанием по разнообразию и количеству в инфраклассе Pteriomorpha. Показано также, что белок, кодируемый второй открытой рамкой считывания, является неотъемлемым структурным компонентом практически всех полноразмерных элементов *L31*. Полученные данные способствуют лучшему пониманию эволюции представителей *ITm*-транспозонов. Дальнейшее изучение *L31*-транспозонов в других таксонах (стрекающие), а также исследование функции белка, кодируемого второй рамкой считывания, позволит лучше понять эволюцию ДНК-транспозонов, механизмы их горизонтального переноса и вклад в биоразнообразие эукариот.

Ключевые слова: ДНК-транспозоны, *L31*-транспозоны, двустворчатые моллюски, эволюция геномов, биоразнообразие

DOI: 10.31857/S0026898424010051, EDN: OFVSWF

ВВЕДЕНИЕ

Мобильные генетические элементы (МГЭ) эукариот — это подвижные элементы, которые существуют и эволюционируют внутри геномов хозяев, перемещаясь из локуса в локус, а также способны к горизонтальному переносу между геномами. МГЭ широко распространены в геномах прокариот и эукариот [1–4]. После вторжения в геном хозяина МГЭ способны увеличивать число своих копий, зачастую очень значительно. Из-за способности к транспозиции и частой амплификации транспозоны являются основными детерминантами размера генома [1, 2, 5, 6]. Активные перемещения и амплификация МГЭ способны изменять первичную структуру ДНК, влиять на работу генов и их функцию, вмешиваться

в процессы регуляции транскрипции, вызывать хромосомные aberrации в геноме хозяина [7–9]. Любопытно, что, внедряясь в геном хозяина, МГЭ не просто хаотично перемещаются, как считалось ранее, а способны “выбирать” наиболее предпочтительные для себя локусы генома [10]. В геноме хозяина МГЭ проходят определенные стадии жизненного цикла, результатами которого, как правило, являются постепенная деградация и элиминация [11–13]. Однако некоторые МГЭ могут существовать в геноме хозяина достаточно долго при большом количестве функциональных копий, например ретротранспозоны *LINE-1* млекопитающих [14, 15]. Кроме того, МГЭ могут избежать исчезновения, став источником новых генов или других полезных для генома структур [2]. На данный

момент выявлен целый ряд структурных и регуляторных генов, возникших в результате молекулярного одомашнивания МГЭ [16]. Также жизненный цикл МГЭ можно перезапустить посредством горизонтального переноса в геном другого хозяина (явление горизонтального переноса ряда элементов, например SINE, не описано) [2, 3, 11].

Другой важный аспект — это способность МГЭ отвечать активностью на некоторые стрессовые физические, химические и биологические факторы [17–22]. Индукция активности МГЭ выражается в увеличении частоты перемещений и, соответственно, усилении амплификации. Все это, в свою очередь, дестабилизирует геном хозяина, приводя к повышению частоты мутаций и гибели организма, а в редких случаях — к их адаптации и дальнейшей эволюции [7, 23, 24]. В соответствии с классификацией МГЭ эукариот делят на два класса: класс I — ретротранспозоны; класс II — ДНК-транспозоны. Деление на классы опирается на механизм пе-

ремещения, свойственный каждому из них. Ретротранспозоны перемещаются через создание РНК-посредника. ДНК-транспозоны не создают себе посредника и напрямую вырезаются из геномной последовательности и встраиваются в нее [25–27].

ДНК-транспозоны — очень разнообразная и многочисленная группа МГЭ, включающая три подкласса и не менее 17 надсемейств [26, 28]. Одной из широко распространенных групп ДНК-транспозонов эукариот является инфракласс *ITm* [29, 30], включающий несколько надсемейств: *pogo*, *Tc1/mariner*, *Gambol*, *Sailor* [29, 31–34]. Автономные транспозоны *ITm* содержат, как правило, одну открытую рамку считывания (ОРС), которая кодирует фермент транспозазу и окружена концевыми инвертированными повторами (КИП). Встречаются элементы, имеющие более одной пары инвертированных повторов. Дополнительные повторы обычно называют субконцевыми инвертированными повторами (СИП) (рис. 1а) [27, 32].

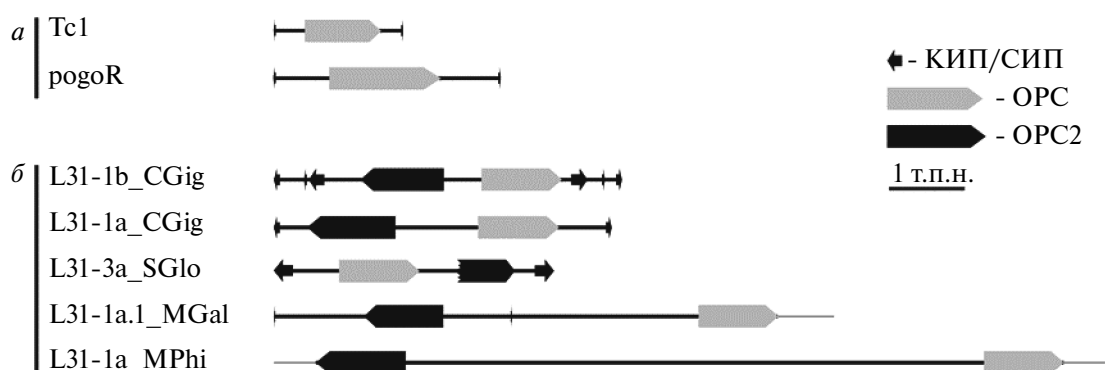


Рис. 1. Структура транспозонов суперсемейств *Tc1/mariner* и *pogo* (а) и разнообразие структуры *L31*-элементов двустворчатых моллюсков (б). КИП/СИП — концевые инвертированные повторы или субконцевые инвертированные повторы; ОРС — открытая рамка считывания, кодирующая транспозазу; ОРС2 — открытая рамка считывания, кодирующая белок с неизвестной функцией.

В ходе исследования разнообразия элементов *ITm* в геноме тихоокеанской устрицы *Crassostrea gigas* был описан транспозон *Mariner-31_CGi* — представитель неизвестной обособленной группы, условно названной *L31* (от like *Mariner-31_CGi*) [35]. На филогенетическом дереве данная группа заняла место вне известных крупных надсемейств *Tc1/mariner* и *pogo*. ОРС представителей *L31* кодировала транспозазу с каталитическим доменом DD37E. При этом в последовательности элемента *Mariner-31_CGi* присутствовала еще одна ОРС, кодирующая белок с неизвестной функцией. Детальное изучение представителей данной группы не входило в задачи работы, поэтому исследователи ограничились только анализом распространенности элементов *L31* среди эукариот. На основании данных о наличии гомологий с транспозазой *Mariner-31_CGi*

представители *L31* были найдены только у двух таксонов: двустворчатых моллюсков (*Bivalvia*) и стрекающих (*Cnidaria*) [35].

Позднее в работе, посвященной глубокому анализу филогенетических отношений и классификации элементов *ITm*, представители *L31* были отнесены к так называемым минорным группам [30]. При этом высказывались сомнения в принадлежности элемента *Mariner-31_CGi* к *L31*, который, собственно, и был референсным для данной группы. Кроме того, изучение структурных особенностей и эволюции элементов *L31* не входило в задачи работы. Таким образом, группа *L31* оставалась практически неизученной.

В данной работе представлен первый детальный анализ элементов *L31* у двустворча-

тых моллюсков, в результате которого описаны структура, разнообразие и распространенность, а также установлено филогенетическое положение транспозонов *L31* в инфраклассе *ITm*. Показано также, что белок, кодируемый *OPC2*, является неотъемлемым структурным компонентом практически всех полноразмерных элементов *L31*.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Поиск элементов. В качестве образца для поиска элементов *L31* использовали объединенную аминокислотную последовательность транспозазы и белка, кодируемого *OPC2* (Дополнительные материалы 1), принадлежащих описанному ранее элементу *Mariner-31_CGi* устрицы *Crassostrea gigas* [35], взятые из базы данных Repbase (<https://www.girinst.org/>). Полногеномные нуклеотидные последовательности представителей двустворчатых получены из базы данных NCBI Assembly (<https://www.ncbi.nlm.nih.gov/assembly/>). Поиск полноразмерных элементов *L31* осуществляли с помощью tBLASTn [36] по гомологии с образцом, представленным в полногеномных последовательностях двустворчатых. При анализе обнаруженных гомологий мы учитывали последовательности с процентом соответствия образцу по длине (Query Coverage) 45 и более. Далее из оставшегося многообразия мы выбирали все последовательности, кодирующие потенциально функциональную транспозазу (с доменом DD37E) и потенциально функциональный белок *OPC2*. Критериями для оценки потенциальной функциональности транспозаз и белков *OPC2* были: неповрежденная *OPC*, наличие стартового кодона и стоп-кодона, а для транспозаз еще и длина, составляющая не менее 340 аминокислотных остатков (а.о.). Если таковых не было, то выбирали наиболее сохранившиеся последовательности. В тех случаях, когда последовательность, гомологичная образцу, кодировала потенциально функциональный белок *OPC2*, но не кодировала потенциально функциональную транспозазу, мы также анализировали ее как предполагаемый элемент *L31*. За допустимое расстояние между белком *OPC2* и транспозазой внутри предполагаемого элемента мы приняли 10 000 п.н. При большем расстоянии белок *OPC2* и транспозазу считали принадлежащими разным элементам или разным копиям одного элемента.

При поиске копий потенциально функциональных транспозаз для каждого обнаруженного элемента в качестве образца брали аминокислотную последовательность неповрежденной транспозазы этого элемента и осуществляли поиск в tBLASTn с настройками фильтров: Query Coverage 99-100 и Percent Identity 90-100. Сре-

ди результатов поиска при подсчете учитывали только те транспозазы, идентичность которых составляет не менее 90%, а соответствие образцу по длине не менее 99% (Query Coverage), которые имеют такой же домен, как у образца (например, DNLSAH), а также неповрежденную *OPC* со стартовым кодоном и стоп-кодоном. При поиске потенциально функциональных копий каждого обнаруженного белка *OPC2* мы брали в качестве образца последовательность этого белка. Поиск копий осуществляли в tBLASTn с настройками фильтров: Query Coverage 99-100, при подсчете учитывали только копии без внутренних стоп-кодонов.

КИП обнаруживали с помощью BLASTn путем анализа нуклеотидных последовательностей [36]. Границы гипотетических *OPC* определяли с помощью ORF Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>) и далее уточняли визуально.

Филогенетический анализ. Для филогенетического анализа были взяты *L31*-транспозазы двустворчатых моллюсков, описанные в данном исследовании, а также *L31*-транспозазы тихоокеанской устрицы *C. gigas* (*Mariner-31_CGi* и *Mariner-53_CGi*) из Repbase и транспозазы элементов, представляющих известные группы инфракласса *ITm* (табл. 1, см. Дополнительные материалы http://www.molecbio.ru/downloads/2024/1/supp_Puzakov_rus.pdf), в общей сложности 139 аминокислотных последовательностей. Множественное выравнивание проводили с использованием MAFFT с применением метода G-INS-I [37]. Филогенетическое дерево получено с использованием метода максимального правдоподобия в программе IQ-TREE [38] со сверхбыстрым бутстреп-анализом (UFBoot) (1000 повторов) [39], модель LG+I+G4 выбрана с помощью ModelFinder [40].

Анализ доменной структуры транспозазы. В анализ доменной структуры последовательностей взяты все элементы, имеющие по первичному анализу потенциально функциональные транспозазы (табл. 2). Расположение GRPR-подобного мотива и маркерных аминокислотных остатков: аспартат (D), аспартат (D) и глутамат (E) каталитического домена идентифицировали визуально по гомологии. ДНК-связывающий домен (три α -спирали до GRPR-подобного мотива и три после) выявляли, анализируя вторичную структуру транспозазы, предсказанную с помощью программы PSIPRED v4.0 [50]. Предполагаемую последовательность сигнала ядерной локализации (NLS) определяли с помощью программы PSORT (<https://www.genscript.com/psort.html>). Графическое представление обобщенных последовательностей отдельных участков каталитического домена сгенерировано с помощью WebLogo [51].

Таблица 1. *ITm*-транспозоны, транспозазы которых использовали в филогенетическом анализе

<i>Famar1</i>	AAO12863	<i>Guest_Ca-sativa</i>	XP010462775
<i>Dmmar1</i>	AAA28678	<i>Guest_Soymar1</i>	AF078934
<i>Tymar1</i>	AAP45328	<i>Guest_Br-oleracea</i>	XP013589454
<i>Hsmar1</i>	AAC52010	<i>Guest_Phylostachys_edulis</i>	ADP24264
<i>Bytmar1</i>	CAD45367	<i>Guest_Pisum_sativum</i>	AAX51974
<i>Quetzal</i>	AAB02109	<i>Gambol_(AAAB01008815)</i>	AAAB01008815
<i>Mariner-14_CGi</i>	Repbase	<i>Gambol_(AAAB01008960)</i>	AAAB01008960
<i>SsTRT</i>	[41]	<i>Gambol_(AAAB01008968)</i>	AAAB01008968
<i>An-gambiae1</i>	AF378002	<i>Gambol_(AAAB01016702)</i>	AAAB01016702
<i>In_Rhinella_marina</i>	[42]	<i>Gambol_(AAAB01006894)</i>	AAAB01006894
<i>DD35E_TR-Xihe</i>	[43]	<i>Gambol_(AAAB01006919)</i>	AAAB01006919
<i>DD38E_IT_At</i>	[44]	<i>Gambol_(AAAB01008879)</i>	AAAB01008879
<i>TLEWI-1_BPl</i>	[45]	<i>Gambol_(AAAB01008849)</i>	AAAB01008849
<i>VS-Maze</i>	[46]	<i>Gambol_(AAAB01008958)</i>	AAAB01008958
<i>L18-1_HVul</i>	[47]	<i>TBE_AAA18578</i>	[30]
<i>Z-1_POch</i>		<i>TBE_AAB42017</i>	
<i>Bmmar1</i>	[48]	<i>TBE_AAB42032</i>	
<i>pogoR11</i>	S20478	<i>TBE_AAB49643</i>	
<i>Tigger1</i>	U49973	<i>TBE_AAB49646</i>	
<i>Fot1</i>	Q00832	<i>TBE_AAB58026</i>	
<i>Tan1</i>	U58946	<i>TBE_AAB58028</i>	
<i>Pot2</i>	Z33638	<i>TBE_AAB58030</i>	
<i>pogo-5_PBac</i>	[49]	<i>TBE_AAB58032</i>	
<i>pogo-2_BOva</i>		<i>TBE_AAB58034</i>	
<i>Tec_AAA62601</i>	[30]	<i>TBE_AAB58036</i>	
<i>Tec_AAA91339</i>		<i>TBE_AAB58377</i>	
<i>Tec_AAM80490</i>		<i>TBE_TBE1</i>	
<i>Tec_Tec1</i>		<i>TBE_AAB42034</i>	
<i>Sailor_Mo_Teggra</i>	[29]	<i>TBE_EJY78953</i>	
<i>Sailor_Mo_Batpla</i>		<i>TBE_EJY85485</i>	
<i>Sailor_Mo_Cepnem</i>		<i>IS630Ss</i>	X05955
<i>Sailor_Mo_Cragig</i>		<i>IS630Se</i>	NP_073225
<i>Sailor_Mo_Cycsin</i>		<i>IS630_Citrobacter_braakii</i>	STH95988)
<i>Sailor_Mo_Hallae</i>		<i>IS630_Escherichia_coli</i>	GDW80866)
<i>Sailor_Mo_Halrub</i>		<i>IS630_Shigella_dysenteriae</i>	VDG84061)
<i>Sailor_Mo_Halruf</i>		<i>IS630_Shigella_flexneri</i>	SRR10263
<i>Sailor_Mo_Limfor</i>		<i>HvSm_XP_004209659</i>	[30]
<i>Sailor_Mo_Mermer</i>		<i>HvSm_XP_004212365</i>	
<i>Sailor_Mo_Modphi</i>		<i>HvSm_XP_012557766</i>	
<i>Sailor_Mo_Mytcor</i>		<i>HvSm_M-6_SM</i>	Repbase
<i>Sailor_Mo_Pinimb</i>		<i>L31_Mariner-53_CGi</i>	
<i>Sailor_Mo_Rudphi</i>		<i>L31_Mariner-31_CGi</i>	
<i>Sailor_Mo_Sacglo</i>			

Предсказание функции белка, кодируемого второй ОРС. Гипотетическую функцию белка, кодируемого второй ОРС, изучали с использованием в качестве референсной аминокислотной последовательности элемента *L31-1b_CGig* (*C. gigas*) протяженностью 368 а.о. Поиск гомологий с известными консервативными доменами осуществляли с помощью сервиса CDD [52], алгоритма BLAST [53] и программы Dompred [54]. Кроме того, использовали инструменты предсказания функции FFPred 3 [55] и MEMSAT [56] и сервиса предсказания третичной структуры SWISS-MODEL [57].

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Эволюционное разнообразие L31-транспозонов двустворчатых

В результате поиска гомологов *Mariner-31_CGi* у двустворчатых моллюсков обнаружены 45 уникальных элементов (табл. 2). Для моделирования эволюционных связей проведен филогенетический анализ, в который вошли *L31*-транспозоны двустворчатых, а также представители всех известных групп инфракласса *ITm*. Полученная дендрограмма (рис. 2) была укоренена на элементы *IS630*, которые, как правило, используются в качестве внешней группы при филогенетическом анализе транспозонов *ITm* [25–30, 35].

Установлено, что *L31*-транспозоны формируют единую кладу с высокой достоверностью (бутстреп-значение 100%) (рис. 2). Высокую значимость в сформированных кладах показали и представители надсемейств *Sailor* и *pogo*, а также так называемых минорных групп *Tec* и *TBE*. Элементы *HvSm* создали единую смешанную ветвь с элементами надсемейства *Gambol* со значимостью 99%. В исследовании, в котором элементы *HvSm* выделены в отдельную кладу, элементы *Gambol* не вошли в филогенетический анализ [30], поэтому есть вероятность, что элементы *HvSm* и *Gambol* являются представителями одной эволюционной группы. Представители надсемейства *Tc1/mariner* сформировали кладу с достоверностью 69% (рис. 2), что может указывать на более высокую филогенетическую гетерогенность данной группы или на раннюю дивергенцию элементов внутри группы.

Внутри клады *L31* наблюдалось достаточно высокое разнообразие. Мы выделили пять кластеров с достоверностью от 70 до 100% (рис. 2). В кластерах встречалось более одного элемента, обнаруженного в одной геномной сборке, поэтому в названиях элементов к номерам, соответствующим кластеру, мы добавляли литеры (рис. 2). Каждый кластер отличался по входя-

щим в него представителям отрядов двустворчатых моллюсков. Так, в кластер *L31-1* вошли элементы отрядов *Mytilida* и *Ostreida* с доминирующим преобладанием разнообразия последнего. Десять элементов *Ostreida* распределились в три ветви (a, b и c) и один отдельный элемент *Saccostrea glomerata*, получивший литеру d (*L31-1d_SGlo*), так как не вошел ни в одну из трех ветвей, упомянутых выше. Также в *L31-1* попали два элемента *C. gigas* из Repbase: *Mariner-31_CGi* (описанный ранее [35]) и *Mariner-53_CGi* (не охарактеризованный). Шесть элементов *Mytilida* объединились в единую группу (рис. 2). Кластер *L31-2* сформировали элементы представителей отрядов *Ostreida*, *Mytilida* и *Pectinida*. Здесь наблюдалось преобладание элементов, обнаруженных в геномах *Mytilida*, которые распределились в три ветви. В *L31-3* вошли транспозоны отрядов *Ostreida*, *Mytilida* и *Pterioda*, а в *L31-4* – элементы отрядов *Mytilida*, *Pterioda*, *Adapedonta* и *Venerida*. Немногочисленность элементов в кластерах не позволяет выделить явные клады, хотя все же наблюдается группировка элементов в соответствии с таксономией. Кластер *L31-5* включает только два элемента, которые выявлены у представителей отряда *Pectinida* (рис. 2).

Распространение L31-транспозонов среди двустворчатых

На момент исследования в коллекциях NCBI были представлены полногеномные последовательности только двустворчатых подкласса *Autobranchia*, поэтому подкласс *Protobranchia* остался неизученным. В результате анализа распространности элементов *L31* среди двустворчатых подкласса *Autobranchia* установлено, что они преобладают в инфраклассе *Pteriomorphia* (рис. 3). В другом инфраклассе – *Heteroconchia* – только у трех из 17 видов обнаружено по одному достаточно хорошо сохранившемуся *L31*-транспозону (табл. 2). Все три вида принадлежат к субтерклассу *Euheterodonta*: два из отряда *Venerida* и один из *Adaendonta*. Все элементы принадлежат к кластеру *L31-4*. Еще у одного представителя отряда *Venerida* (*Ruditapes philippinarum*) обнаружены короткие фрагменты, гомологичные *Mariner-31_CGi*. У других видов, принадлежащих отрядам *Venerida* (три вида), *Adapedonta* (один вид), *Mytilida* (три вида) и *Cardiida* (два вида) субтеркласса *Euheterodonta*, гомологий не выявлено (табл. 2). В другом субтерклассе *Palaeoheterodonta* инфракласса *Heteroconchia* анализ геномных последовательностей четырех представителей отряда *Unionida* выявил в трех из них короткие фрагменты *L31*-транспозонов (табл. 2).

Таблица 2. Элементы L31 двустворчатых моллюсков

Отряд	Вид/ Идентификатор геном- ной сборки	Элемент	Дли- на, п.н. ^a	КИП, п.н. ^a	СИП, п.н. ^a	Транспозаза, а.о. ^a	Блок ОРС2, а.о. ^a	Интервал между ОРС, п.н. ^a	Число копий	Число ПФТ	Число ПФБ-2
Ostreida	<i>Crassostrea ariakensis</i> GCA_020458035	<i>L31-1a_CAr</i>	4631	85/85	—	359	390	1194	8	2	4
		<i>L31-1c_CAr</i>	4704	34/34	—	356	385	992	4	3	3
		<i>L31-1a_CGig</i>	4450	85/85	—	359	390	1086	4	1	2
	<i>Crassostrea gigas</i> GCA_011032805	<i>L31-1b_CGig</i>	4597	78/76	36/35 209/224	355	368	485	8	2	1
		<i>L31-2a.1_CGig</i>	3686	145/145	—	356	219	676	6	4	0
		<i>L31-2a.2_CGig</i>	3860	145/145	—	356	188	483			
	<i>Crassostrea virginica</i> GCA_002022765	<i>L31-1a_CVir</i>	4127	129/129	—	356	395	719	21	9	5
		<i>L31-1a.1_CHon</i>	4481	87/87	—	355	387	1106	6	4	4
		<i>L31-1a.2_CHon</i>	4279	86/86	—	359	390	913			
	<i>Crassostrea hongkongensis</i> GCA_015776775	<i>L31-1b_CHon</i>	3682	217/203	—	355	368	487	25	2	2
		<i>L31-1c.1_CHon</i>	2958	31/31	—	356	—	—	10	0	0
		<i>L31-1c.2_CHon</i>	2912	—	—	364	239	1097			
<i>L31-1c.3_CHon</i>		2823	—	—	207	383	1047				
<i>L31-2a_CHon</i>		3841	145/145	—	357	126	816	16	2	0	
<i>Saccostrea glomerata</i> GCA_003671525	<i>L31-1c_SGlo</i>	3400	—	—	361	332	1972	2	0	0	
	<i>L31-1d_SGlo</i>	1738	—	—	347	117	343	5	1	0	
	<i>L31-2a_SGlo</i>	5552	—	—	354	293	3605	3	0	0	
	<i>L31-3a_SGlo</i>	3670	255/254	—	356	252	503	4	0	0	
	<i>L31-1b_OLur</i>	7144	—	—	335	285	5186	1	0	0	
<i>Mytilus coruscus</i> GCA_017311375	<i>L31-1a_MCor</i>	5859	50/49	—	97	420	2463	2	0	0	
	<i>L31-2a_MCor</i>	1041	—	—	346	—	—	2	0	0	
	<i>L31-1a_MEdu</i>	4707	22/22**	—	157	370	3126	2	0	0	
	<i>L31-2a_MEdu</i>	6282	31/31	—	358	217	1412	10	3	0	
	<i>L31-2b_MEdu</i>	6357	213/214	—	355	253	2494	8	5	0	
<i>Mytilus galloprovincialis</i> GCA_900618805	<i>L31-1a.1_MGal</i>	5477	23/23**	—	355	355	3341	2	1	1	
	<i>L31-1a.2_MGal</i>	1278	—	—	—	425	—				
	<i>L31-2a_MGal</i>	10181	28/28	—	347	205	1416	6	0	0	
	<i>L31-2b_MGal</i>	9975	77/77	—	355	252	2585	8	1	0	
<i>Gigantidas platifrons</i> GCA_002080005	<i>L31-3a_GPla</i>	13921	—	—	361	101	12527	5	0	0	
	<i>L31-1a_MCal</i>	7339	32/32	34/36*	352	431	2004	1	0	0	
	<i>L31-2a.1_MCal</i>	5385	30/30	—	358	246	1177	9	2	0	
	<i>L31-2a.2_MCal</i>	5739	27/27	—	252	219	1102				
<i>L31-2a.3_MCal</i>	590	—	—	195	—	—					

Таблица 2. Продолжение

Отряд	Вид/ Идентификатор геном- ной сборки	Элемент	Дли- на, п.н. ^а	КИП, п.н. ^а	СИП, п.н. ^а	Транспозаза, а.о. ^а	Блок ОРС2, а.о. ^а	Интервал между ОРС, п.н. ^а	Число копий	Число ПФТ	Число ПФБ-2
	<i>Modiolus philippinarum</i> GCA_002080025	<i>L31-1a_MPhi</i>	9887	—	—	353	397	7631	3	1	1
		<i>L31-1b_MPhi</i>	4658	—	—	322	291	2563	2	0	0
		<i>L31-2a_MPhi</i>	1074	—	—	357	—	—	3	1	0
		<i>L31-2b_MPhi</i>	2207	—	—	363	220	455	4	1	0
		<i>L31-3a_MPhi</i>	6184	104/104	—	355	169	3342	3	0	0
		<i>L31-3b_MPhi</i>	2445	—	—	358	169	861	5	0	0
		<i>L31-4a_MPhi</i>	3403	—	—	357	307	1405	2	1	0
	<i>Perna viridis</i> GCA_GCA_018327765	Короткие фрагменты									
Pectinida	<i>Limnoperna fortunei</i> GCA_GCA_003130415	Короткие фрагменты									
	<i>Pecten maximus</i> GCA_020750765	<i>L31-2a_PMax</i>	3784	27/27	—	356	215	794	9	4	0
	<i>Argopecten irradians</i> GCA_004382745	<i>L31-2a_AIrr</i>	2343	—	—	359	149	816	6	1	0
	<i>Mizuhopecten yessoensis</i> GCF_002113885	<i>L31-5_AIrr</i>	1038	—	—	345	—	—	3	2	—
		<i>L31-5_MYes</i>	969	—	—	323	—	—	6	0	—
Pterioida	<i>Pinctada imbricata</i> GCA_002216045	<i>L31-3a_PImb</i>	2532	31/31	—	357	—	—	3	2	—
		<i>L31-3b.1_PImb</i>	3908	367/360	—	154	147	1565	2	0	0
		<i>L31-3b.2_PImb</i>	4433	37/37	342/338	294	196	1067			
		<i>L31-4a.1_PImb</i>	4456	47/47	—	357	456	428	24	22	20
		<i>L31-4a.2_PImb</i>	2559	—	—	357	352	426			
		<i>L31-4b_PImb</i>	10108	165/166	—	359	327	4941	6	1	0
		<i>L31-4c.1_PImb</i>	6923	19/19	197/204	356	—	—	10	0	1
		<i>L31-4c.2_PImb</i>	4760	237/267	52/52	273	459	490			
Arcoida	<i>Pinna nobilis</i> GCA_016161895	Нет сходства									
	<i>Anadara kagoshimensis</i> GCA_021292105	Короткие фрагменты									
	<i>Tegillarca granosa</i> GCA_013375625	Короткие фрагменты									
Venerida	<i>Mercenaria mercenaria</i> GCF_014805675	<i>L31-4a_MMer</i>	15077	71/71	306/312	348	207	9218	13	9	0
	<i>Cyclina sinensis</i> GCA_012932295	<i>L31-4a_CSin</i>	7678	30/30	—	344	212	5515	8	6	0
	<i>Archivesica marissinica</i> GCA_014843695	Нет сходства									

Таблица 2. Окончание

Отряд	Вид/ Идентификатор геном- ной сборки	Элемент	Дли- на, п. н. ^а	КИП, п. н. ^а	СИП, п. н. ^а	Транспозаза, а.о. ^а	Белок ОРС2, а.о. ^а	Интервал между ОРС, п. н. ^а	Число копий	Число ПФТ	Число ПФБ-2
Adapedonta	Ruditapes philippinarum GCA_009026015	Короткие фрагменты									
	Lutaria rhynchaena GCA_008271625	Нет сходства									
	Corbicula fluminea GCA_001632725	Нет сходства									
	Solen grandis GCA_021229015	L31-4a_SGra	959	–	–	314	–	–	1	0	0
Myida	Raponea genetosa GCA_902825435	Нет сходства									
	Mya arenaria GCA_922144925	Нет сходства									
	Dreissena polymorpha GCA_020536995 GCA_000806325	Нет сходства									
	Dreissena rostriformis GCA_007657795	Нет сходства									
Cardiida	Gari tellinella GCA_922989275 GCA_922984925	Нет сходства									
	Sinonovacula constricta GCA_009762815 GCA_007844125	Нет сходства									
	Potamilus streckersoni GCA_016746295	Короткие фрагменты									
Unionida	Megaloniais nervosa GCA_016617855	Нет сходства									
	Margaritifera margaritifera GCA_015947965	Короткие фрагменты									
	Venustaconcha ellipsiformis GCA_003401595	Короткие фрагменты									

^а Приведены данные для репрезентативной копии.
* Инвертированные повторы фланкируют только ген транспозазы.
** Инвертированные повторы фланкируют только ОРС2.
Условные обозначения: КИП – концевые инвертированные повторы, СИП – субконцевые инвертированные повторы, а.о. – аминокислотный остаток, ОРС – открытая рамка считывания, ПФТ – потенциально функциональная транспозаза; ПФБ-2 – потенциально функциональный белок, кодируемый ОРС2.

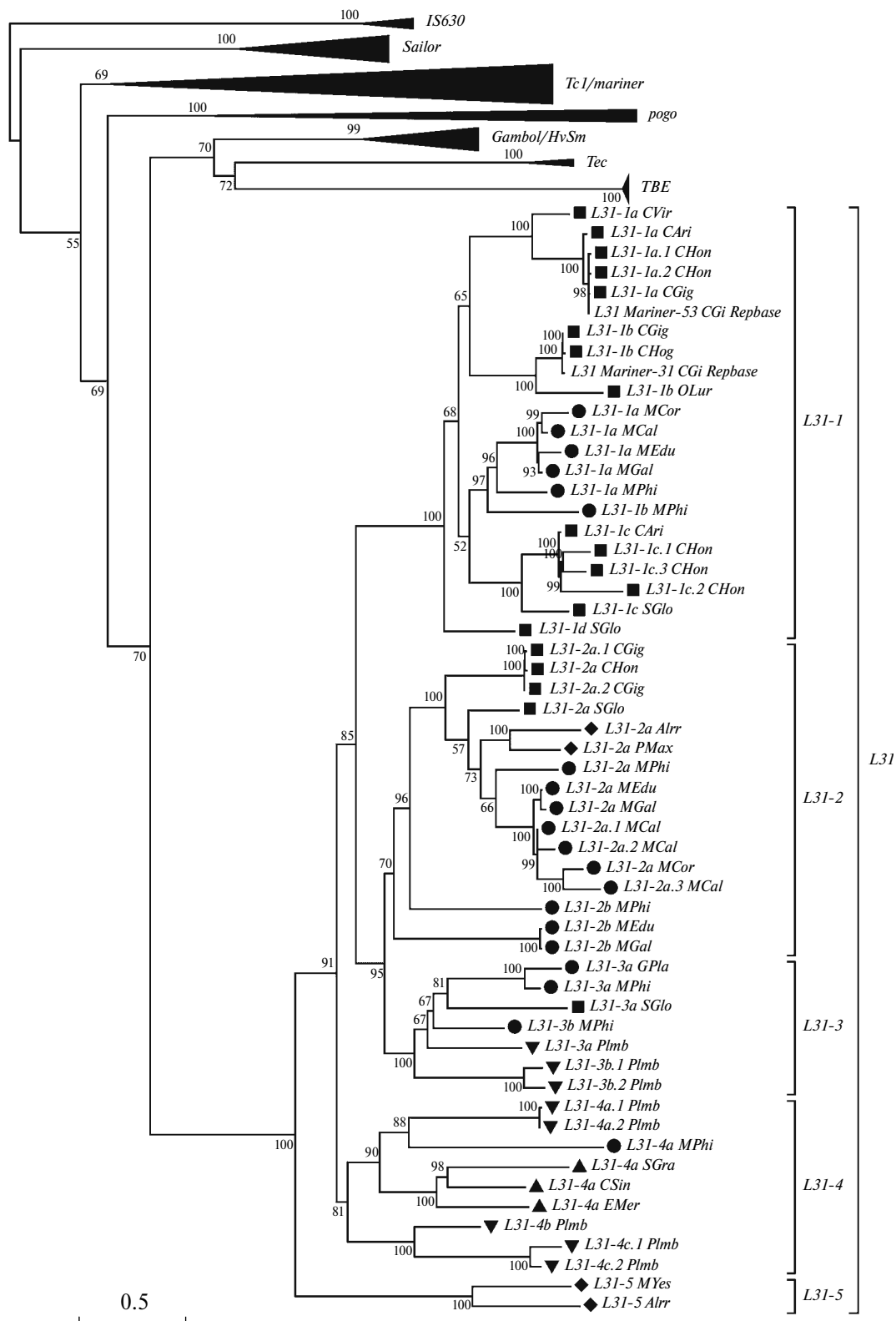


Рис. 2. Филогенетическое разнообразие *L31*-элементов двусторчатых моллюсков. Названия клад указаны справа от дендрограммы. Геометрическими фигурами обозначены транспозоны, выявленные в данном исследовании: квадрат — отряд Ostreida, круг — отряд Mytilida, ромб — отряд Pectinida, треугольник вершиной вверх — отряды Adapedonta и Venerida, треугольник вершиной вниз — отряд Pterioidea. Бутстреп-значения менее 50% на дендрограмме не указаны.

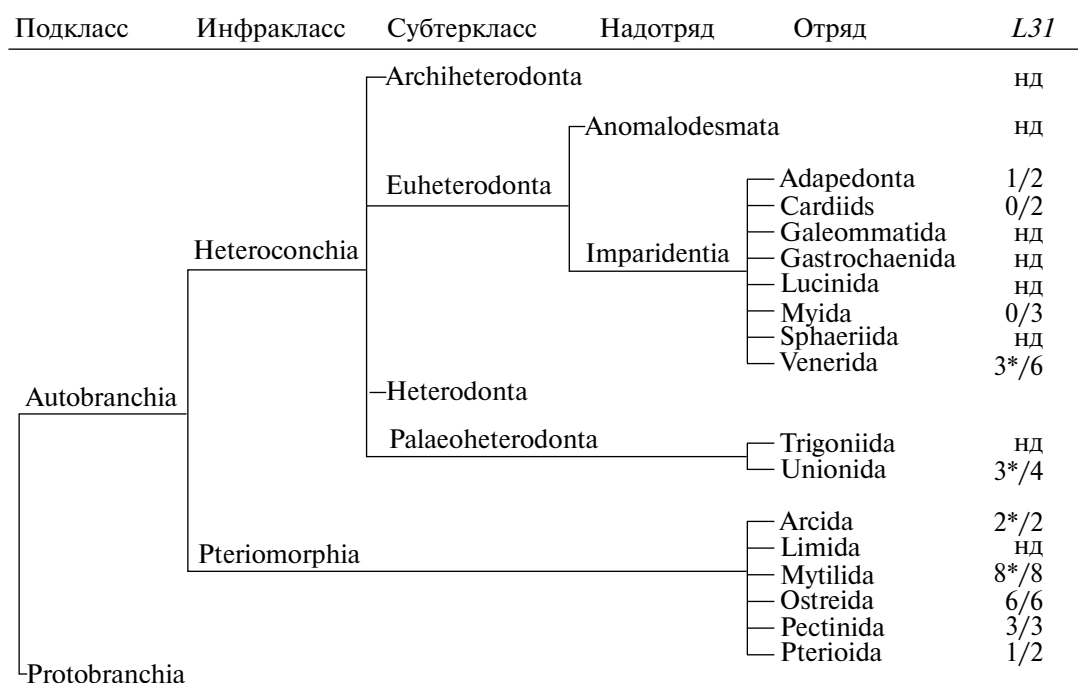


Рис. 3. Распространение L31-транспозонов среди двустворчатых. нд — нет данных; косая черта разделяет количество видов, у которых обнаружены L31-элементы, и общее число исследованных видов таксона; *у некоторых видов выявлены только короткие фрагменты (обрывки) L31-элементов (см. текст и табл. 2). Таксономическое дерево создано на основе данных из World Register of Marine Species (WoRMS) (<https://www.marinespecies.org/>) и TimeTree (<http://www.timetree.org/>).

Элементы L31 выявлены в четырех (Mytilida, Ostreida, Pectinida, Pterioida) из шести отрядов инфракласса Pteriomorpha. У двух видов отряда Arcoida обнаружены только короткие фрагменты, гомологичные *Mariner-31_CGi*. Полногеномные сборки у представителей Limoida в коллекциях NCBI на момент исследования отсутствовали (табл. 2).

В геномах всех шести изученных видов отряда Ostreida обнаружены элементы L31. При этом наблюдалась вариабельность в представленности транспозонов в геномах. Количество уникальных элементов L31 варьировало от 1 до 4 (табл. 2). Представители кластера L31-1 выявлены во всех изученных сборках, тогда как элементы L31-2 найдены у трех организмов, а L31-3 — только у одного.

У Mytilida L31-транспозоны обнаружены у шести из восьми видов. У двух видов (*Limnoperna fortunei* и *Perna viridis*) выявлены только короткие фрагменты элементов L31. Вариабельность по количеству уникальных элементов была еще более высокой, чем у Ostreida (от 1 до 7) (табл. 2). У представителей рода *Mytilus* (четыре вида) выявлены элементы кластеров L31-1 и L31-2, тогда как у *Modiolus philippinarum* обнаружены представители четырех кластеров (все, кроме L31-5), а у *Gigantidas platifrons* только элемент L31-3 (табл. 2, табл. 3).

Во всех трех изученных геномах представителей Pectinida присутствовали 1 или 2 элемента L31 (табл. 2, табл. 3). У гребешка *Pecten maximus* — элемент кластера L31-2, а у *Mizuhopecten yessoensis* — L31-5. В геноме *Argopecten irradians* обнаружены представители обеих этих групп (L31-2 и L31-5).

В отряде Pterioida анализировали геномные сборки двух видов (табл. 2). Однако L31-транспозоны выявлены только у *Pinctada imbricata*. При этом наблюдалось высокое разнообразие: пять уникальных элементов представляли кластеры L31-3 и L31-4 (табл. 2, табл. 3).

Особенности L31-транспозонов двустворчатых

Наиболее яркой чертой элементов L31 является дополнительная ОРС (ОРС2). Показано, что преобладающее большинство элементов, как и *Mariner-31_CGi* (*C. gigas*), содержат ОРС2. При этом во всех элементах (за исключением L31-3a_SGlo) ОРС и ОРС2 направлены от центра к краям транспозона, тогда как у L31-3a_SGlo они направлены в одну сторону (рис. 1б). Таким образом, ОРС2 является характерным компонентом элементов L31, а не случайным спутником *Mariner-31_CGi*, как предполагалось ранее [35].

Таблица 3. Характеристика кластеров *L31*-транспозонов двустворчатых

Кластер	Длина ^А , п.н.	КИП, п.н.	СИП, п.н.	Транспо- заза, а.о.	Белок ОРС2, а.о.	Интервал между ОРС, п.н.	Число копий	Число ПФТ	Число ПФБ-2	Отряд
<i>L31-1</i>	2958–9887	31–203	35–224	322–364	332–431	343–7631	1–25	0–9	0–4	Ostreida Mytilida
<i>L31-2</i>	3686–10181	27–214	–	345–363	–	455–3605	2–16	0–5	0	Ostreida Mytilida Pectinida
<i>L31-3</i>	2532–13921	31–360	338–342	355–361	–	503–12527	2–5	0–2	0	Ostreida Mytilida Pterioidea
<i>L31-4</i>	4456–15077	19–267	52–312	344–359	327–459	426–9218	1–24	0–22	0–20	Mytilida Pterioidea Adapedonta Venerida
<i>L31-5</i>	969–1038 ^Б	–	–	323–345	–	–	3–6	0–2	0	Pectinida

Примечание. КИП – концевые инвертированные повторы; СИП – субконцевые инвертированные повторы; а.о. – аминокислотный остаток; ОРС – открытая рамка считывания; ПФТ – потенциально функциональная транспозаза; ПФБ-2 – потенциально функциональный белок, кодируемый второй ОРС; А – минимальное значение, указано только для элементов с КИП; Б – у данного кластера нет элементов с КИП, поэтому указана протяженность фрагментов, гомологичных транспозазе.

Наряду с эволюционным разнообразием транспозаз, элементы *L31* также различаются общей протяженностью, длиной КИП и субконцевых инвертированных повторов (СИП) как между кластерами, так и внутри них (табл. 3). Наименьшая протяженность элемента, имеющего КИП (*L31-3a_PImb*), составляла 2532 п.н., но этот элемент не имел ОРС2. Самый короткий элемент, имеющий КИП и обе ОРС, – *L31-3a_SGlo*, длина которого составляет 3670 п.н. При этом встречаются элементы, длина которых превышает 10 000 п.н. и даже достигает 15 077 п.н. (*L31-4a_MMer*). Во всех кластерах, кроме *L31-5*, представлены как короткие, так и длинные варианты. У обоих представителей *L31-5* не сохранились КИП и ОРС2, выявлены лишь последовательности, гомологичные транспозазе (табл. 2, табл. 3).

Флуктуации в протяженности элементов *L31* преимущественно обусловлены длиной межгенного интервала и в меньшей степени длиной КИП, а также наличием и длиной СИП. Во всех кластерах (с обеими ОРС) наблюдается значительный разброс длины межгенного интервала (табл. 3) – от нескольких сотен до нескольких тысяч пар нуклеотидов. Однако у преобладающего числа элементов протяженность этого интервала варьирует в диапазоне от 400 до 2500 п.н., только у отдельных представителей его длина достигает более значительных величин (например, 9218 п.н. у *L31-4a_MMer* или 12 527 п.н.

у *L31-3a_GPla*). Причиной такого разнообразия в длине межгенного интервала могут быть мутационные процессы, сопровождающие эволюционный путь *L31*-транспозонов, в результате которых возникают как делеции, так и вставки (иногда протяженные) фрагментов ДНК.

КИП и СИП также демонстрируют разнообразие – от коротких вариантов 19 п.н. (*L31-4c.1_PImb*) до длинных 367 п.н. (*L31-3b.1_PImb*). Обобщая полученные данные, можно отметить отсутствие СИП в кластере *L31-2*, при том что он имеет достаточно много представителей (13 элементов). Также есть случаи, когда СИП фланкируют только ген транспозазы (*L31-1a_MCal*) или единственная пара инвертированных повторов фланкирует только ОРС2 (*L31-1a_MEdu*, *L31-1a.1_MGal*) (табл. 2, рис. 1б).

Подсчет копий каждого уникального элемента в геномах двустворчатых показал, что *L31*-транспозоны не имели высокой транскрипционной активности в прошлом. На это указывает немногочисленность сохранившихся экземпляров (табл. 2). Из 45 обнаруженных элементов только три (*L31-1b_CHon*, *L31-1a_CVir*, *L31-4a_PImb*) имели более 20 копий, еще у двух (*L31-4a_MMer*, *L31-2a_CHon*) было 13 и 16 копий соответственно. Число копий у остальных *L31*-транспозонов не превышало 10, а у трех элементов (*L31-1b_OLur*, *L31-1a_MCal*, *L31-4a_SGra*) выявлено лишь по одной копии.

Между кластерами больших различий не установлено. В группах *L31-3* и *L31-5* отсутствуют элементы с числом копий, превышающим 10, но и в остальных группах присутствие более многокопийных элементов носит скорее эпизодический характер.

Около половины обнаруженных *L31*-транспозонов (27 элементов) сохранили копии с ОРС, кодирующей потенциально функциональную транспозазу, а только около четверти (11 элементов) – потенциально функциональный белок ОРС2 (табл. 2). Количество копий с интактной ОРС транспозазы не превышало 10, а копий с интактной ОРС2 – пяти. Исключением является *L31-4a*, у которого 22 из 24 копий несли интактный ген транспозазы и 20 из них имели потенциально функциональный ген белка ОРС2.

Полноразмерные транспозазы выявлены во всех кластерах надсемейства *L31* и варьировали от 322 до 364 а.о. В каждом кластере найдены элементы как имеющие копии с потенциально функциональной транспозазой, так и без таковых (табл. 2, табл. 3).

Поскольку белок, кодируемый ОРС2, ранее не был описан, то не установлено, какую длину можно считать характерной для него. В связи с этим мы считали полноразмерными все варианты, превышающие 300 а.о. Копии элементов, содержащие полноразмерные белки ОРС2, выявлены только в кластерах *L31-1* и *L31-4*, хотя в кластерах *L31-2* и *L31-3* также обнаружены элементы с последовательностями, гомологичными ОРС2, но кодирующими белок протяженностью менее 300 а.о. В кластерах *L31-1* и *L31-4* копии с потенциально функциональными белками ОРС2 выявлены не у всех элементов, но у некоторых присутствовали в относительно большом количестве (20 копий у *L31-4a_PImb*) (табл. 2). Таким образом, можно отметить, что последовательности ОРС2 в большей степени подвержены деградации и элиминации.

Доменная структура транспозазы *L31*-транспозонов двустворчатых

Основными компонентами функциональных транспозаз элементов инфракласса *ITm* являются ДНК-связывающий и каталитический (DDE/D) домены [32]. ДНК-связывающий домен расположен в первой половине (N-концевой части) аминокислотной последовательности транспозазы и выявляется по наличию шести α -спиралей. Этот домен обеспечивает связывание транспозазы с КИП. Между первой и второй триадами α -спиралей располагается GRPR-подобный мотив. Этот компонент обеспечивает взаимодействие ДНК-связывающего домена с сайтом-мишенью (динуклеотид ТА) [58]. Вто-

рая половина (С-концевая часть) транспозазы содержит DDE/D-домен, который обладает эндонуклеазной и лигирующей активностью, необходимой для вырезания и вставки транспозона. Название (DDE/D) домена основано на присутствии триады консервативных маркерных аминокислотных остатков – два аспартата (D) и третий либо глутамат (E), либо аспартат. Между первым и вторым маркерными аспартатами находятся, как правило, от 90 до 110 а.о. Расстояние между вторым и третьим маркерными остатками (составляет обычно от 30 до 40 а.о.) является консервативным и часто используется как классификационный признак транспозонов *ITm* (например, семейство *Visitor* – DD40-41D, семейство *mariner* – DD34D, семейство *Tc1* – DD34E) [32–34]. Также транспозазы *ITm* могут содержать сигнал ядерной локализации (NLS), который, как предполагается, способствует транспорту транспозазы из цитоплазмы в ядро [59, 60].

Изучение структуры полноразмерных транспозаз у элементов *L31* двустворчатых показало, что ДНК-связывающий домен, GRPR-подобный мотив и каталитический домен присутствуют практически во всех последовательностях (рис. 4). В отдельных случаях некоторые структуры не обнаружены. Так, в элементе *L31-1a_MPhi* не найдена третья α -спираль, а в *L31-1a_MGal* – шестая α -спираль. В целом вторая триада α -спиралей ДНК-связывающего домена имеет большую вариабельность, чем первая триада α -спиралей. GRPR-подобный мотив в четырех последовательностях был неизвестен (*L31-1a_MPhi*, *L31-1c_Cari*, *L31-5_AIrr*, *L31-4a_MMer*) (рис. 4). NLS-мотив найден в четырех *L31*-транспозонах (*L31-2a_MEdu*, *L31-2a_MCal*, *L31-4a_MPhi*, *L31-4a_CSin*) (рис. 4). Каталитический домен сохранился во всех полноразмерных транспозазах и имел паттерн DD37E. Только элемент *L31-5_AIrr* содержал DD38E домен (рис. 4).

Для сравнения каталитических доменов элементов *L31* разных кластеров мы выбрали три области, включающие каждый из маркерных аминокислотных остатков триады DDE протяженностью 10 а.о., и создали обобщенные последовательности (рис. 5). Сходства и различия в обобщенных последовательностях между всеми кластерами *L31*-транспозонов подтвердили адекватность подразделения на группы. Кластер *L31-5* включал только два элемента в связи с чем обобщенную последовательность области глутамата (третий маркерный остаток) получить не удалось. Однако области первых двух маркерных остатков были более консервативными и также имели отличия от других кластеров (рис. 5).

L31-2a MEdu	-----MATKKNIYGREYLAGKATGRDMSRLIVQEMIEAGANLKTEGEVPRGVTTIKADFKKINQSVTNFWKRYVSEGTISQKKKEKTMGRKK-LNEPDVRLIEFTIKKNPSITARELKDITLLRYSE
L31-2b MEdu	-----MERQNKYGRFQKQFKLSDDFTLTIIDKCLEYGGNSDNQITIPRGTFSKVSEELKVDTDFVRKMKQFCIDK-EVKCKPHK---GLOPKLSNEDREYILAKMKKPTISLSELRKLLHNSV
L31-1a MGa1	-----MKRNNQGRQISAGKALGDDLRGLIVHELKESGHHVGN-SIPKGIAPKVAEYKIKTQTVHNHKKYNELD-SVSRRCAG---GRPKRYGIDEIEFVNVLKTERPSPQNTLRDQLLQYSA
L31-2b MGa1	-----MERQNKYGRFQKQFKLSDDFTLTIIDKCLEYGGNSDNQITIPRGTFSKVSEELKVDTDFVRKMKQFCIDK-EVKCKPHK---GLOPKLSNEDREYILAKMKKPTISLSELRKLLHNSV
L31-1a MPhi	-----MEKKKISGRSYINGKALGDGLRSLLIDDLADAGAKIGE-KMPKGLVGRVAEYKVCOTTIYRLWKYSEDLD-SIKRRPYST---GRPKKYGEELIEFVHVLKKEPMSMOLTKVKDQLLOYSN
L31-2a MPhi	MESEDSCKSTNDFGRQFQKGRALSLEFRLQIVQEKIKKISGG-SLLEVPFGIYKVGKRFKIHTSSVTIKWRRFLENG-CPSEPRPHGK-QNVKLTIPDDQFQFIEVLKREKPSICIKDIRKELYEHS
L31-4a MPhi	-----MATRKNHSGRECPMGKAIQDQLRLQIVQNLVLESNGNASTGEVPLGTITNISKLIKISQSVNIVWRRFISDA-SVAEKKRTSN-GNPKLTIPEDVRMIEFLKRETTSTAKELKEKLTREYSPV
L31-1b Cgig	-----MEKEKRF---ISQKGVSLSKDMRKLIVTKLKEIGSDEESKFPSPGAFSIVRSNLDLRLRVLGRVWSEFCERR-DFSTLKDKNKGGRITSLSNSDIGYLEFLKTSDDPAISYQKMKNEKLMFSS
L31-2a-1 Cgig	-----MCTKNNYGREYQOGIELSDTFRSLLIQDLIDRGCDASAGKVPWALSATAKKEFKVSVTKVNKIWRRFVEHG-TCTEKMNRKS-GPSKLTQOEDVDHIAVLKTNTPMSLKSVDNLIQYSN
L31-1a Cvir	-----MAKINSAGRVYDKGTPLGIDLRSSIINYMEEQAKFGSLQPLPGLPQVSEIFKVSHPILVTIKWQYQIEG---IVKLPEYKG-GRKRLKDDEEDVQIHFLKTKPSMLKSVEEVLKYSN
L31-1d SGlo	-----MEKRNKYGRVFNKGLSNDLRGLIVQDCEMGG-----VPGKVKELAKKYKLNPDITIRVMLKHIEGE-QISKKPT---GRPKLTDEEDVEFISALKTARPSMOLETIRDOVIONSN
L31-1a CHon	-----MCTKNNYGREYQOGIELSDTFRSLLIQDLIDRGCDASAGKVPWALSATAKKEFKVSVTKVNKIWRRFVEHG-TCTEKMNRKS-GPRKLTQEDVQFIECLKRETPMSYKEIQELKRRYSPV
L31-1a Cari	-----MLSVNRRNIAGRVYKGIPTGIDYRRSIIDYMESNGARLGEYSIPRGLQKTASQOQKVPSPATINFTWQYCNEG---CVKVPVQTNRGRKKKLEEDVEVVRFLKHVRPSMPLQTVRDELLKNSN
L31-1c Cari	-----MESLSKQGRNLINGKSLSRDMNVAICDLVNGGASAGDLKFPGLGANLAKKYKISKSTVSNIRWKYNE---NFSIPRKMNTKDRKIGNEDREFIKALKNERPSITQLOQITRDELLKNSN
L31-1a-1 CHon	-----MLSVNRRNIAGRAYEKIPTGIDYRRSIIDYMESNGARLGEYSIPRGLRKTASQKEKVPATITTSFTWQYCNEG---CVKVPVQTNRGRKKKLEEDVEVVRFLKHVRPSMPLTVRDELLKNSN
L31-2a-1 CHon	-----MLSVNRRNIAGRAYEKIPTGIDYRRSIIDYMESNGARLGEYSIPRGLRKTASQKEKVPATITTSFTWQYCNEG---CVKVPVQTNRGRKKKLEEDVEVVRFLKHVRPSMPLTVRDELLKNSN
L31-1b CHon	-----MCTKNNYGREYQOGIELSDTFRSLLIQDLIDRGCDASAGKVPWALSATAKKEFKVSVTKVNKIWRRFVEHG-TCTEKMNRKS-GPSKLTQOEDVDHIAVLKTNTPMSLKSVDNLIQYSN
L31-1a Cgig	-----MLSVNRRNIAGRAYEKIPTGIDYRRSIIDYMESNGARLGEYSIPRGLRKTASQKEKVPATITTSFTWQYCNEG---CVKVPVQTNRGRKKKLEEDVEVVRFLKHVRPSMPLTVRDELLKNSN
L31-2a Alrr	-----MATRKNKYGRYHNGTGLSDDLRLSLIQEILGNGNSQSGVSPRGITVKVANTRFVDEKSVTRFWKKYISBG-SMSETRARHKA-GNPKLNGADVDQLVSLKNTPSVITGNKIKRNLCKYSS
L31-2a PMax	-----MTTRKNKSGREYKKGDLGSDSLRLKLIQVETLHLGNGVLTEGIPRGVFNKVAEYKVNQRTVSNFWQYVSDG-ALSEKTRNTL-GRPKLNTDIEHIFLKKRPSITGTTELKRLLEYSVP
L31-5 Alrr	-----MNRNDIVNNGRVYKKGDLGSDSLRLKLIQVETLHLGNGVLTEGIPRGVFNKVAEYKVNQRTVSNFWQYVSDG-ALSEKTRNTL-GRPKLNTDIEHIFLKKRPSITGTTELKRLLEYSVP
L31-4b Plmb	-----MKKSHVNNKGRYDVGKPLGEDLSLIVQYVLEKGANSESRFIPRGEMAKAGERFNVSNNTIKNTWLYQDSG-DVKHRQAR-GPRKLTDEEDFNIAAKKKEPFTLENIKKEKLNQA
L31-2a-1 Plmb	-----MSKSKKGLYKFGTGLSADFLRLNISKLKDYGANDDTGSPRGIAKVKSKELLIDKNCVTRAWLSWINDK-NVESKPK---GPTKGSMTSDQNDLHIEFLKREKPSISLEIHNNKL
L31-4a-2 Plmb	-----MSKSKKGLYKFGTGLSADFLRLNISKLKDYGANDDTGSPRGIAKVKSKELLIDKNCVTRAWLSWINDK-NVESKPK---GPTKGSMTSDQNDLHIEFLKREKPSISLEIHNNKL
L31-1a Plmb	-----MATRNMNNGRYHNGTGLSDDLRLSLIQEILGNGNSQSGVSPRGITVKVANTRFVDEKSVTRFWKKYISBG-SMSETRARHKA-GNPKLNGADVDQLVSLKNTPSVITGNKIKRNLCKYSS
L31-2a-1 MCal	-----MATKKNIYGREYLAGKATGRDMSRLIVQEMIEAGANLKTEGEVPRGVTTIKADFKKINQSVTNFWKRYVSEGTISQKKKEKTMGRKK-LNEPDVRLIEFTIKKNPSITARELKDITLLRYSE
L31-4a MMer	-----MATKKSKPKG-----GRPLSDDFNLILKMINFGADENLNIKPRGAVTSVARQELIDRGTVSNIRWKYCATK-NCAPSPYIR---HASKGLTDEHKQYIIFELPQTPSLSLGSIKOLKAMC
L31-4a CSin	-----MKRLEK-----KILSEDFLLVIRIKKILGGDTETNTVPRGTATVARSCGVHQTIVKNKIWDKYCLTG-EVKVMKPNKG---GRKKTFCTEEMFNLYLVKQKPSISLGLKDKLSQC
L31-2a MEdu	A-SANV-DVSTIYRTMSRDL---DFTFKR---LHRPSGDRFTPR-NMRYTQAYLDFQCQTKRPHQIKFDESGFKLVTA-NRNYGHSKK-GEQCIEIGRFIEGANLTLNCLIGLDCVLYNFVGDGSPNS
L31-1a MGa1	-----ISSI-STSTVSRIITNDL---NMTYKRTIHYKKNRPMVR-NLQYQOFLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-2b MGa1	VPNNNV-SRMTICRVIKKDL---NMTYK---VSRPKAERTFAA-NLRYTQAFIDHITQDLANKLIFWDESGYVVVTA-NRTGHSV-GTRCVEREYHPNPVNLTLVLGNGRMHYNFVDGTS
L31-2b MPhi	-----IEAI-SYASISRMINDL---NMTYK---VTHYKKEFTLG-NLQYQOFLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-2a MPhi	I-DGGV-SLSTIGRAVNRNLLGKGWTLKK---LERPLGDRFSIP-NLRYTQAYLDFIDFQOVRHKLFDDESGFKTLTA-NKTRGHSK-GRRCIEGVPFRDNPNTLNLVLGDCVLYNFVGDGSPNS
L31-2a MPhi	-----TNNV-HEITYYRAINRKL---NMTYK---LIRPIGRDRFSIAYIMRYQAFIDYDQCTQPHQIKFDESGFKLVTA-NRNYGHSK-GRPCIEIRKIHVGPNTLNLVLGDCVLYNFVGDGSPNS
L31-4a MPhi	-----TTNV-SVQKQRAVKGKFLGGEWTKYK---LTTFSKDRFTDG-NIEYTESFEMELLHRADEPYKLFKFDDESGFCTPDVGTPKYGHALK-GTRGVMMNPNRKTMTTDLVLGLTGVMAHKTFRGVS
L31-1b Cgig	T-LQNI-SLPTVSNIRKDL---HMTLKR---VTFCGNRFTLP-NLQYQOFLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-2a-1 Cgig	-----SGEV-SLPTINRCILKDL---KTYK---LKYSRIFERTND-NMYSQAYIDFQCTQNTQIKFDESGFKLVTA-NRNYGHSK-GRPCIEIRKIHVGPNTLNLVLGDCVLYNFVGDGSPNS
L31-1a Cvir	A-VIKV-SESTISRHKNDL---NMTYK---IARYSKNRFTTPQ-NMNYTQFRLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-1d SGlo	-----VQNV-SLPTVSNIRKDL---NMTYK---ITKYDKDRFTLP-NLQYQOFLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-1a Cvir	-----SGEV-SLPTINRCILKDL---NMTYK---LKYSRIFERTND-NMYSQAYIDFQCTQNTQIKFDESGFKLVTA-NRNYGHSK-GRPCIEIRKIHVGPNTLNLVLGDCVLYNFVGDGSPNS
L31-1a Cari	S-METV-GLSTISRTLKEDL---NMTYK---ISKINKNRFTTPQ-NMNYTQFRLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-1c Cari	-----TISI-SISTISNTIRKDL---NMTYK---ISCLYKNRFTLP-NLQYQOFLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-1a-1 CHon	S-IEITL-GLSTISRTLKEDL---NMTYK---ISKINKNRFTTPQ-NMNYTQFRLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-2a-1 CHon	S-IEITL-GLSTISRTLKEDL---NMTYK---ISKINKNRFTTPQ-NMNYTQFRLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-1b CHon	T-LQNI-SLPTVSNIRKDL---HMTLKR---VTFCGNRFTLP-NLQYQOFLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-1a Cgig	S-IEITL-GLSTISRTLKEDL---NMTYK---ISKINKNRFTTPQ-NMNYTQFRLMYVSNKDFPFLKFDMEGMVGLVDG-QPVYGHSGK-GPCVEITRYDPHANTASLIIGITGVKMYIIEGASDS
L31-2a Alrr	-----SGNV-HNSTIYRAMSKL---DLTYK---INRPSCDRFTQY-NMNYTQAFIDFQCTQNTQIKFDESGFKLTT-NPAGYHSGK-GQACVEIKGRVGRNLTNLVLGDCVLYNFVGDGSPNS
L31-2a-1 Alrr	-----SGKV-DSTIYRTMSRDL---DLTYK---ITRPSCDRFTQY-NMNYTQAFIDFQCTQNTQIKFDESGFKLTT-NPAGYHSGK-GQACVEIKGRVGRNLTNLVLGDCVLYNFVGDGSPNS
L31-5 Alrr	CTAENVPSKISINKVRQVYL---DMTFKK---ISQIPEETTRPN-HEDTVTRIAKMTNTPQMHFDEASVISTSG-NRSGYHSTI-GKAKYAEVORASNTKTPVNMCCGYFLDYLVDGSPNS
L31-4b Plmb	-----NKEV-TMTTISALRKLIGQWTRKV---LTKATERFTPA-NEQYTEAIEQSKPLRDEAGFALSAEAVNRKTHAPK-LARIEAOKMRKTNTLNLVLGDCVLYNFVGDGSPNS
L31-4a-1 Plmb	ENCNKIIHESITSHALKQNLAAEFTTRK---LCKPARERTSD-NLKYTEALMYLSQKDVHRIKYDEAGFNSRDC-SPIYGHALR-GERAVEVSSKAKSNTLNLMLGVNGVCHNVGDGATD
L31-4a-2 Plmb	ENCNKIIHESITSHALKQNLAAEFTTRK---LCKPARERTSD-NLKYTEALMYLSQKDVHRIKYDEAGFNSRDC-SPIYGHALR-GERAVEVSSKAKSNTLNLMLGVNGVCHNVGDGATD
L31-3a Plmb	-----TGRV-DEQLIGKAVRKLPSGENTRKI---ITRONRNYTDG-NMAYTQWLYNIQKQVNRKILKFDDESGFKITDA-NKRYGHSPI-GEKIEIGKYVATPNTLNLMLGVNGVCHNVGDGATD
L31-2a-1 MCal	-----NANV-DVSTIYRTMSRDL---DFTFKR---LHRPSGDRFTTP-NMRYTQAYLDFQCQTKRPHQIKFDESGFKLVTA-NRNYGHSK-GEQCIEIGRFIEGANLTLNCLIGLDCVLYNFVGDGSPNS
L31-4a MMer	-----NVDI-SKAPICGFLK---EMTRKI---LTKPAERFND-NMAYTRAFVLAFLDTPDKMKFDESGFKHDPVDCNRVGRSTK-GERAIEITPKANTNTNLNLIGLDCVLYNFVGDGSPNS
L31-4a CSin	-----DLEV-STSTVSRIITNDL---YNTKRL---LTPRAADRFTDI-NRYTQAFIDALHQDQVASFIFDESGFALPDVANPRYGRSPR-GDRAIEITHDKRKIPNKTNLMLIGINGVMFANILDGSPNS
L31-2a MEdu	ERYLNFWHEASLSDQCYGRPTFLPGDILLVNDCAIHNNQS-ERILNTFFNMGGIDYGLFVPSYDPLNP-IEICFSKIKTVIKQERFKELVSNKLKLV---IKAIQEINQSDIHGFYRHTGYFNV
L31-1a MGa1	GEYLQFGEASQSYNTDGEVFPQGECLVVDNAPTHNNMS-ERVLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-2b MGa1	NTYLFQMGESAHNTENGSIYIAPGDTIIVDNSPLHNNRA-EVTLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-1a MPhi	VEFLQFGEAQSQYTDGEPVFTPGQCLVVDNAPTHNNLS-ERVLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-2b MPhi	AFYVNFHEAIESVTDEGVPFLSPGDIIVVDNAPTHNNDS-EVLNWFRLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-2a MPhi	ATYLNFWHEASLSDQIFGRPLFLPGDILLVNDCAIHNNQA-EQVLHAFHFGMGVEYAFLLPTTSPDLPNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-4a MPhi	FDFMNFITECVHSYVQFGLALQPGDILLVNDCAIHNNSEV-ARSTKRWLEAGQIDVFTPRYSPDLPNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-1b Cgig	IEFLHFLGEGNAVTEDEGERVLQGRDLSIVDNPAPTHNNMS-ERVLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-2a-1 Cgig	YTYLNFWHEAQQNDYEGRAVLIPGDFVVDNCPHNNNS-ERVLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-1a Cvir	TEYLFQFGEAQSQYTDGEPVFPNGDCLVVDNAPTHNNMS-ERVLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-1d SGlo	YTYLNFWHEAQQNDYEGRAVLIPGDFVVDNCPHNNNS-ERVLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-1a CHon	YTYLNFWHEAQQNDYEGRAVLIPGDFVVDNCPHNNNS-ERVLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-1a Cari	YTYLNFWHEAQQNDYEGRAVLIPGDFVVDNCPHNNNS-ERVLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-1c Cari	TKYDVFGEAANSFTDDGERAIQGVGVVDNAPTHHAA-ERILNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-1a-1 CHon	TEYVQFIAEATQSDTDEGVPFPHGDVLIADNAAIHNNRA-ERELNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-2a-1 CHon	TEYVQFIAEATQSDTDEGVPFPHGDVLIADNAAIHNNRA-ERELNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-1b Cgig	TEYVQFIAEATQSDTDEGVPFPHGDVLIADNAAIHNNRA-ERELNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-1a Cgig	TEYVQFIAEATQSDTDEGVPFPHGDVLIADNAAIHNNRA-ERELNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-2a Alrr	NTYLNFWHEASLSDQNMGCPLFKPGDGVVDNCPHNNHA-ERVLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-2a PMax	MEYLNFWHEASLSDQNLGRVFMAGDGVVDNCPHNNHA-ERVLNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-5 Alrr	DEFLKIDFQASDSDGSLVVDNAPTHHAA-ERILNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-4b Plmb	DEFLKIDFQASDSDGSLVVDNAPTHHAA-ERILNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-4a-1 Plmb	QYLNFWFQDAISDTEGHALVPGDGVVDNAPTHHAA-ERILNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-4a-2 Plmb	QYLNFWFQDAISDTEGHALVPGDGVVDNAPTHHAA-ERILNWLFTTGVMEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-3a Plmb	DNFVEYFVFAAESYTSIGMPVLEPGDGVVDNCPHNNHKEG-ERRAKDFLFGMGIEYFLTPAYSPLDNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-2a-1 MCal	ERYLNFWHEASLSDQCYGRPTFLPGDILLVNDCAIHNNQS-ERILNTFFNMGGIDYGLFVPSYDPLNP-IEICFSKIKTVIKQERFKELVSNKLKLV---IKAIQEINQSDIHGFYRHTGYFNV
L31-4a MMer	DTMFEVFNALNSTTNGDALRPLGDFLIVDNCPIHRRFA-EDVLSRMLDRGIEYIILPTTSPHLPNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN
L31-4a CSin	DTVNFVFLQAINVANNAGDFALRPGDILLVNDCAIHNNRA-EELIKLDFYRHLGIEYIIPATYSPLHNPABCFRKYKLLKSDRFGPVLQRDLKQVAV---YKAFNEITLMDTSTFFKATGEYMN

Рис. 4. Множественное выравнивание последовательностей транспозаз *L31*-транспозонов. α -Спираль ДНК-связывающего домена выделены серым. Предполагаемый NLS обозначен полужирным курсивом. Триада DDE каталитического домена выделена черным. GPRK-мотив обозначен полужирным и подчеркнут.

Группа элементов	Область первого маркерного аминокислотного основания DDE-домена	Область второго маркерного аминокислотного основания DDE-домена	Область третьего маркерного аминокислотного основания DDE-домена
<i>L31-1</i>	LKFMDEMGTK	TVDNAPTHHN	SPDLNPAEAC
<i>L31-2</i>	TKFMDESGEK	TVDNCAITHHN	SPDLNPVENC
<i>L31-3</i>	LKFFDESGEK	TVDNCPTHKN	SPDLNPAENC
<i>L31-4</i>	LKFEDEAGFA	VVDNAPTHRE	SPDLNPVELC
<i>L31-5</i>	LMFFDEASTT	VEDNAGFEHQ	RDELNECENE

Рис. 5. Особенности консервативных районов каталитического домена транспозазы элементов *L31*. Маркерные аминокислотные остатки каталитического домена выделены серым.

Предполагаемая функция белка, кодируемого второй OPC *L31*-транспозонов

Поскольку в ходе исследования выяснилось, что белок, кодируемый OPC2, является постоянным и достаточно консервативным компонентом *L31*-транспозонов, мы решили подробнее изучить его предполагаемые функции. В качестве референсной использовали последовательность белка, кодируемого OPC2 элемента *L31-1b_CGig* (*C. gigas*), протяженностью 368 а.о.

Несмотря на то что ранее в аминокислотной последовательности, кодируемой OPC2 элемента *Mariner-31 CGi*, был выявлен домен Myosin_tail_1 (pfam01576) [35], поиск гомологий с известными консервативными доменами в коллекциях NCBI с помощью BLAST в этот раз не дал результатов. Анализ с помощью сервисов поиска доменов CDD и Dompred также не выявил никаких гомологий.

Анализ с помощью FFPred 3, предназначенным для предсказания функции (в терминах генной онтологии) аминокислотных последовательностей, когда гомология с известными белками малоинформативна, позволил выявить несколько возможных вариантов. Среди функций, связанных с биологическими процессами, наибольшие значения достоверности имели транспорт (GO:0006810), регуляция транскрипции на основе нуклеиновых кислот (GO:1903506) и регуляция биосинтеза РНК (GO:2001141). Среди предсказанных молекулярных функций наиболее достоверными были: связывание с белками цитоскелета (GO:0008092), связывание с нуклеиновыми кислотами (GO:0003676) и связывание с актином (GO:0003779). Наиболее вероятными структурами из компонентов клеток, с которыми связана предполагаемая функция белка, кодируемого OPC2, были мито-

хондрии (GO:0005739), митохондриальная мембрана (GO:0031966) и мембрана (GO:0016020).

Исследование с помощью MEMSAT, который позволяет на основе аминокислотных последовательностей предсказать топологию, показало, что кодируемый OPC2 белок может быть трансмембранным. При этом N-конец белка (1–37 а.о.) — это предположительно его внеклеточная часть, 38–53 а.о. — трансмембранная, а с 54 а.о. до конца — внутриклеточная. Эти данные отчасти согласуются с предсказанием связи с мембранами, полученными в результате анализа с использованием FFPred 3. α -Спиральные трансмембранные белки участвуют в клеточной передаче сигналов, транспорте мембрано-непроницаемых молекул, межклеточной связи, распознавании клеток и клеточной адгезии [52].

Анализ белка, кодируемого OPC2 элемента *L31-1b_CGig*, с помощью сервиса предсказания третичной структуры на основе поиска гомологий с известными белками SWISS-MODEL выявил сходство с четырьмя белками: белком комплекса ядерной поры (нуклеопорин) NUP58 (5ijn.1.G), субъединицей E фактора транскрипции II (TFIIE) (5oqm.1.U), белком с цинковыми пальцами NBR1 (2bkf.1.A) и тектином-2 (7rro.32.A). При этом центральная часть белка, кодируемого OPC2, гомологична NUP58, TFIIE и тектину-2, тогда как N-конец сходен только с NBR1.

Нуклеопорин NUP58 входит в состав комплекса ядерной поры, который обеспечивает транспорт молекул через ядерную мембрану в обоих направлениях [61]. Субъединица E фактора транскрипции II участвует в плавлении ДНК в области промотора во время транскрипции [62]. Тектин-2 является компонентом центриолей [63], а цинковые пальцы являются

модулями, взаимодействующими с ДНК, РНК, другими белками или небольшими молекулами [64]. Обобщая функциональные аспекты, с которыми связаны четыре приведенных белка, можно выделить трансмембранный транспорт, участие в синтезе РНК, взаимодействие с ДНК, РНК и белками, что коррелирует с данными, полученными с использованием FFPred 3.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В результате исследования *L31*-транспозонов двустворчатых моллюсков получена детальная информация о разнообразии, распространении и структуре этих элементов. Эти данные позволяют классифицировать *L31*-транспозоны как самостоятельное надсемейство, входящее в большую группу (инфракласс) *ITm*. Внутри надсемейства нами выделено пять кластеров, что свидетельствует о его эволюционном разнообразии. Из всех моллюсков *L31*-транспозоны выявлены только у двустворчатых [35]. Однако и внутри этого класса наблюдается ограниченное распространение элементов *L31*. На данный момент показано присутствие *L31*-транспозонов в подклассе *Autobranchia* с преобладанием по разнообразию и количеству в инфраклассе *Pteriomorphia* (рис. 3). Ограниченное распространение дает основание предполагать, что предковый *L31*-транспозон двустворчатых появился после дивергенции этой группы организмов. Класс двустворчатые отделился от брюхоногих приблизительно 530 млн лет назад [65]. Предковый *L31*-транспозон мог возникнуть у моллюсков как в ходе эволюции элементов *ITm*, так и проникнуть в прародителя таксона в результате горизонтального переноса. Явление горизонтального переноса играет значимую роль в широком распространении и эволюции ДНК-транспозонов [66, 67]. Функциональные, транспозиционно-активные элементы эукариот способны колонизировать геномы новых хозяев и мультиплицироваться, внося вклад в эволюцию и биоразнообразие. Описано довольно много случаев горизонтального переноса элементов инфракласса *ITm* [67].

Наиболее вероятно, что эволюционное и структурное разнообразие *L31*-транспозонов двустворчатых является следствием эволюции представителей этого надсемейства ДНК-транспозонов уже внутри таксона. Это связано с этапом диверсификации, который в ходе “жизненного цикла” транспозонов следует, как правило, после колонизации генома хозяина. “Жизненный цикл” включает этапы колонизации, диверсификации, деградации, элиминации [11]. Альтернативными итогами окончания “жизненного цикла” ДНК-транспозонов могут быть молекулярная доместикация, горизонтальный

перенос в новый геном и рестарт “жизненного цикла” внутри прежнего генома [11, 35, 48, 68]. Распространение представителей различных кластеров *L31*-транспозонов среди отрядов двустворчатых указывает на то, что событие диверсификации также произошло достаточно давно, так как в одном отряде могут встречаться элементы разных кластеров (табл. 3).

Малое количество копий отражает, по-видимому, невысокую транспозиционную активность *L31*-транспозонов. В сочетании с высоким разнообразием это может свидетельствовать о возможных событиях так называемого рестарта “жизненного цикла” внутри прежнего генома, когда в результате мутационных процессов появляется активный транспозон с функциональной транспозазой. Относительно высокое число копий с потенциально функциональной транспозазой у элемента *L31-4a_PImb* (*Pinctada imbricata*) может свидетельствовать о том, что он по-прежнему активен или был активным сравнительно недавно (по эволюционным меркам).

Вариации в длине элементов и размерах КИП (СИП), на наш взгляд, также отражают долгий эволюционный путь представителей надсемейства *L31*. Однако несмотря на древность происхождения, у многих элементов сохранилась вторая ОРС, что дает основание предполагать необходимость продукта этого гена для транспозиционной активности *L31*-транспозонов. Близкими эволюционными группами надсемейства *L31* являются *TBE* и *Tec* [30], что подтверждается и в данной работе (рис. 2). Примечательно, что и *TBE*, и *Tec* по структуре тоже отличаются от основного массива элементов инфракласса *ITm*. Элементы семейства *TBE* имеют небольшие КИП (около 80 п.н.) и несут три ОРС, кодирующих транспозазу, небольшую ОРС с неизвестной функцией и белок с цинковыми пальцами [69, 70]. Элементы *Tec* имеют очень длинные КИП (около 700 п.н.) и три ОРС: транспозазы, белка с неизвестной функцией и сайт-специфической рекомбиназы, которая может выполнять транспозицию в отсутствие специальной транспозазы [71, 72]. Представители обеих групп пока выявлены только у инфузорий [69, 71]. Белок ОРС2 *L31*-транспозонов, как и дополнительные белки элементов *TBE* и *Tec*, сохраняется в ходе эволюции (и отбора), поэтому, соответственно, его функция может быть связана с обеспечением транспозиционной активности. Поскольку часто белки обладают мультифункциональностью, то и белок ОРС2 может участвовать как в регуляции транскрипции транспозазы, так и обеспечении транспорта транспозазы через ядерный поровый комплекс, или даже транспорта *L31*-транспозона через клеточную мембрану.

Полученные данные о структуре, разнообразии и распространении элементов *L31* двусторчатых моллюсков способствуют лучшему пониманию эволюции представителей инфракласса *ITm*. Дальнейшее изучение *L31*-транспозонов в других таксонах (стрекающие), а также исследование функции белка второй ОРС позволит лучше понять эволюцию ДНК-транспозонов, механизмы горизонтального переноса и вклад в биоразнообразие эукариот.

Аминокислотные последовательности транспозаз представлены в Дополнительных материалах. (см на сайте http://www.molecbio.ru/downloads/2024/1/supp_Puzakov_rus.pdf)

Работа выполнена в рамках государственного задания ФГБУН ИМБИ «Функциональные, метаболические и токсикологические аспекты существования гидробионтов и их популяций в биотопах с различным физико-химическим режимом» (номер гос. регистрации 121041400077-1).

Работа выполнена без привлечения людей и животных в качестве объектов исследования.

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Arkhipova I.R., Yushenova I.A. (2019) Giant transposons in eukaryotes: is bigger better? *Genome Biol. Evol.* **11**, 906–918. doi: 10.1093/gbe/evz041
2. Bourque G., Burns K.H., Gehring M., Gorbunova V., Seluanov A., Hammell M., Imbeault M., Izsvák Z., Levin H.L., Macfarlan T.S., Mager D.L., Feschotte C. (2018) Ten things you should know about transposable elements. *Genome Biol.* **19**, 199. doi: 10.1186/s13059-018-1577-z
3. Kidwell M.G., Lisch D.R. (2000) Transposable elements and host genome evolution. *Trends Ecol. Evol.* **15**, 95–99. doi: 10.1016/s0169-5347(99)01817-0
4. Sotero-Caio C.G., Platt R.N., Suh A., Ray D.A. (2017) Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* **9**, 161–177. doi: 10.1093/gbe/evw264
5. Gao B., Shen D., Xue S., Chen C., Cui H., Song C. (2016) The contribution of transposable elements to size variations between four teleost genomes. *Mob. DNA.* **7**, 4. doi: 10.1186/s13100-016-0059-7
6. Petrov D.A. (2001) Evolution of genome size: new approaches to an old problem. *Trends Genet.* **17**, 23–28. doi: 10.1016/s0168-9525(00)02157-0
7. Юрченко Н.Н., Коваленко Л.В., Захаров И.К. (2011) Мобильные генетические элементы: нестабильность генов и геномов. *Вавил. журн. генетики и селекции.* **15**, 261–270.
8. Grabundzija I., Messing S.A., Thomas J., Cosby R.L., Bilic I., Miskey C., Gogol-Döring A., Kapitonov V., Diem T., Dalda A., Jurka J., Pritham E.J., Dyda F., Izsvák Z., Ivics Z. (2016) A *Helitron* transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nat. Commun.* **7**, 10716. doi: 10.1038/ncomms10716
9. Craig N.L., Chandler M., Gellert M., Lambowitz A., Rice P.A., Sandmeyer S. (2015) *Mobile DNA III*. Washington, USA: ASM Press.
10. Sultana T., Zamborlini A., Cristofari G., Lesage P. (2017) Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat. Rev. Genet.* **18**, 292–308. doi: 10.1038/nrg.2017.7
11. Blumenstiel J.P. (2019) Birth, school, work, death, and resurrection: the life stages and dynamics of transposable element proliferation. *Genes (Basel)*. **10**, 336. doi: 10.3390/genes10050336
12. Bowen N.J., Jordan I.K. (2007) Exaptation of protein coding sequences from transposable elements. *Genome Dyn.* **3**, 147–162.
13. Venner S., Feschotte C., Biémont C. (2009) Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet.* **25**, 317–323.
14. Boissinot S., Chevret P., Furano A.V. (2000) *L1 (LINE-1)* retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.* **17**, 915–928. doi: 10.1093/oxfordjournals.molbev.a026372
15. Platt R.N. 2nd, Vandeweghe M.W., Ray D.A. (2018) Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res.* **26**, 25–43. doi: 10.1007/s10577-017-9570-z
16. Sinzelle L., Izsvák Z., Ivics Z. (2009) Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cell. Mol. Life Sci.* **66**, 1073–1093. doi: 10.1007/s00018-009-8376-3
17. Chow K.C., Tung W.L. (2000) Magnetic field exposure stimulates transposition through the induction of *DnaK/J* synthesis. *Biochem. Biophys. Res. Commun.* **270**, 745–748. doi: 10.1006/bbrc.2000.2496
18. Бубенщикова Е.В., Антоненко О.В., Васильева Л.А., Ратнер В.А. (2002) Индукция транспозиций *МГЭ 412* отдельно тепловым и холодным шоком в сперматогенезе у самцов дрозофилы. *Генетика.* **38**, 46–55.
19. Del Re B., Garoia F., Mesirca P., Agostini C., Bersani F., Giorgi G. (2003) Extremely low frequency magnetic fields affect transposition activity in *Escherichia coli*. *Radiat. Environ. Biophys.* **42**, 113–118. doi: 10.1007/s00411-003-0192-9
20. Захаренко Л.П., Коваленко Л.В., Перепелкина М.П., Захаров И.К. (2006) Влияние γ -радиации на индукцию транспозиций *hobo*-элемента у *Drosophila melanogaster*. *Генетика.* **42**, 763–767.
21. Васильева Л.А., Выхристюк О.В., Антоненко О.В., Захаров И.К. (2007) Индукция транспозиций мобильных генетических элементов в геноме *Drosophila melanogaster* различными

- стрессовыми факторами. *Информацион. Вестн. ВОГуС*. **11**, 662–671.
22. Чересиз С.В., Юрченко Н.Н., Иванников А.В., Захаров И.К. (2008) Мобильные элементы и стресс. *Информацион. Вестн. ВОГуС*. **12**, 217–242.
 23. Piacentini L., Fanti L., Specchia V., Bozzetti M.P., Berloco M., Palumbo G., Pimpinelli S. (2014) Transposons, environmental changes, and heritable induced phenotypic variability. *Chromosoma*. **123**, 345–354. doi: 10.1007/s00412-014-0464-y
 24. Auvinet J., Graça P., Belkadi L., Petit L., Bonnivard E., Dettai A., Detrich W.H. 3rd, Ozouf-Costaz C., Higuier D. (2018) Mobilization of retrotransposons as a cause of chromosomal diversification and rapid speciation: the case for the Antarctic teleost genus *Trematomus*. *BMC Genomics*. **19**, 339. doi: 10.1186/s12864-018-4714-x
 25. Kojima K.K. (2020) Structural and sequence diversity of eukaryotic transposable elements. *Genes Genet. Syst.* **94**, 233–252. doi: 10.1266/ggs.18-00024
 26. Kapitonov V.V., Jurka J. (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* **9**, 411–412. doi: 10.1038/nrg2165-c1
 27. Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P., Schulman A.H. (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982. doi: 10.1038/nrg2165
 28. Yuan Y.W., Wessler S.R. (2011) The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc. Natl. Acad. Sci. USA*. **108**, 7884–7889. doi: 10.1073/pnas.110420810829
 29. Shi S., Puzakov M., Guan Z., Xiang K., Diaby M., Wang Y., Wang S., Song C., Gao B. (2021) Prokaryotic and eukaryotic horizontal transfer of *Sailor* (dd82e), a new superfamily of *IS630-Tc1-Mariner* DNA-transposons. *Biology (Basel)*. **10**, 1005. doi: 10.3390/biology10101005
 30. Dupeyron M., Baril T., Bass C., Hayward A. (2020) Phylogenetic analysis of the *Tc1/mariner* superfamily reveals the unexplored diversity of *pogo*-like elements. *Mob. DNA*. **11**, 21. doi: 10.1186/s13100-020-00212-0
 31. Shao H.G., Tu Z.J. (2001) Expanding the diversity of the *IS630-Tc1-mariner* superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. *Genetics*. **159**, 1103–1115. doi: 10.1093/genetics/159.3.1103
 32. Tellier M., Bouuaert C.C., Chalmers R. (2015) *Mariner* and the *ITm* superfamily of transposons. *Microbiol. Spectr.* **3**, MDNA3-0033-2014. doi: 10.1128/microbiolspec.MDNA3-0033-2014
 33. Gao B., Wang Y.L., Diaby M., Zong W., Shen D., Wang S., Chen C., Wang X., Song C. (2020) Evolution of *pogo*, a separate superfamily of *IS630-Tc1-mariner* transposons, revealing recurrent domestication events in vertebrates. *Mob. DNA*. **11**, 25.
 34. Coy M.R., Tu Z.J. (2010) *Gambol* and *Tc1* are two distinct families of DD34E transposons: analysis of the *Anopheles gambiae* genome expands the diversity of the *IS630-Tc1-mariner* superfamily. *Insect Mol. Biol.* **14**, 537–546. doi: 10.1111/j.1365-2583.2005.00584.x
 35. Puzakov M.V., Puzakova L.V., Cheresiz S.V. (2018) An analysis of *IS630/Tc1/mariner* transposons in the genome of a pacific oyster *Crassostrea gigas*. *J. Mol. Evol.* **86**, 566–580. doi: 10.1007/s00239-018-9868-2
 36. Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402. doi: 10.1093/nar/25.17.3389
 37. Yamada K.D., Tomii K., Katoh K. (2016) Application of the MAFFT sequence alignment program to large data – reexamination of the usefulness of chained guide trees. *Bioinformatics*. **32**, 3246–3251. doi: 10.1093/bioinformatics/btw4122016
 38. Nguyen L.T., Schmidt H.A., von Haeseler A., Minh B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. doi: 10.1093/molbev/msu30039
 39. Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522. doi: 10.1093/molbev/msx281
 40. Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermini L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*. **14**, 587–589. doi: 10.1038/nmeth.4285
 41. Zhang H.H., Li G.Y., Xiong X.M., Han M.J., Zhang X.G., Dai F.Y. (2016) *TRT*, a vertebrate and protozoan *Tc1*-like transposon: current activity and horizontal transfer. *Genome Biol. Evol.* **8**, 2994–3005. doi: 10.1093/gbe/evw213
 42. Sang Y., Gao B., Diaby M., Zong W., Chen C., Shen D., Wang S., Wang Y., Ivics Z., Song C. (2019) *Incomer*, a DD36E family of *Tc1/mariner* transposons newly discovered in animals. *Mob. DNA*. **10**, 45. doi: 10.1186/s13100-019-0188-x
 43. Zong W., Gao B., Diaby M., Shen D., Wang S., Wang Y., Sang Y., Chen C., Wang X., Song C. (2020) *Traveler*, a new DD35E family of *Tc1/mariner* transposons, invaded vertebrates very recently. *Genome Biol. Evol.* **12**, 66–76. doi: 10.1093/gbe/evaa034
 44. Gao B., Zong W., Miskey C., Ullah N., Diaby M., Chen C., Wang X., Ivics Z., Song C. (2020) *Intruder* (DD38E), a recently evolved sibling family of *DD34E/Tc1* transposons in animals. *Mob. DNA*. **11**, 32. doi: 10.1186/s13100-020-00227-7
 45. Puzakov M.V., Puzakova L.V., Cheresiz S.V. (2020) The *Tc1*-like elements with the spliceosomal introns in mollusk genomes. *Mol. Genet. Genomics*. **295**, 621–633. doi: 10.1007/s00438-020-01645-1
 46. Shen D., Gao B., Miskey C., Chen C., Sang Y., Zong W., Wang S., Wang Y., Wang X., Ivics Z., Song C. (2020) Multiple invasions of *Visitor*, a DD41D family of *Tc1/mariner* transposons, throughout the evolution

- of vertebrates. *Genome Biol. Evol.* **12**, 1060–1073. doi: 10.1093/gbe/evaa135
47. Пузаков М.В., Пузакова Л.В. (2022) Распространенность, разнообразие и эволюция ДНК-транспозонов *L18* (DD37E) в геномах стрекающих (Cnidaria). *Молекуляр. биология.* **56**, 476–490. doi: 10.31857/S0026898422030120
 48. Wang S., Diaby M., Puzakov M., Ullah N., Wang Y., Danley P., Chen C., Wang X., Gao B., Song C. (2021) Divergent evolution profiles of DD37D and DD39D families of *Tc1/mariner* transposons in eukaryotes. *Mol. Phylogenet. Evol.* **161**, 107143. doi: 10.1016/j.ympev.2021.10714349
 49. Puzakov M.V., Puzakova L.V., Cheresiz S.V., Sang Y. (2021) The *IS630/Tc1/mariner* transposons in three ctenophore genomes. *Mol. Phylogenet. Evol.* **163**, 107231. doi: 10.1016/j.ympev.2021.107231
 50. Buchan D.W.A., Jones D.T. (2019) The PSIPRED protein analysis workbench: 20 years on. *Nucl. Acids Res.* **47**, 402–407. doi: 10.1093/nar/gkz297
 51. Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190. doi: 10.1101/gr.849004
 52. Marchler-Bauer A., Bo Y., Han L., He J., Lanczycki C.J., Lu S., Chitsaz F., Derbyshire M.K., Geer R.C., Gonzales N.R., Gwadz M., Hurwitz D.I., Lu F., Marchler G.H., Song J.S., Thanki N., Wang Z., Yamashita R.A., Zhang D., Zheng C., Geer L.Y., Bryant S.H. (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucl. Acids Res.* **45**, D200–D203. doi: 10.1093/nar/gkw1129
 53. Boratyn G.M., Schäffer A.A., Agarwala R., Altschul S.F., Lipman D.J., Madden T.L. (2012) Domain enhanced lookup time accelerated BLAST. *Biol. Direct.* **7**, 12. doi: 10.1186/1745-6150-7-12
 54. Bryson K., Cozzetto D., Jones D.T. (2007) Computer-assisted protein domain boundary prediction using the DomPred server. *Curr. Protein Pept. Sci.* **8**, 181–188. doi: 10.2174/138920307780363415
 55. Cozzetto D., Minneci F., Currant H., Jones D.T. (2016) FFPred 3: feature-based function prediction for all Gene Ontology domains. *Sci. Rep.* **6**, 31865. doi: 10.1038/srep31865
 56. Nugent T., Jones D.T. (2009) Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics.* **10**, 159. doi: 10.1186/1471-2105-10-159
 57. Waterhouse A., Bertoni M., Bienert S., Wong G., Chinikar S., Hajivand Z., Mokhayeri H., Nowotny N., Kayedi M.H. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucl. Acids Res.* **46**, W296–W303. doi: 10.1093/nar/gky427
 58. Ivics Z., Izsvák Z. (2015) *Sleeping Beauty* transposition. *Microbiol. Spectr.* **3**, MDNA3-0042-2014. doi: 10.1128/microbiolspec.MDNA3-0042-2014
 59. Ivics Z., Hackett P.B., Plasterk R.H., Izsvák Z. (1997) Molecular reconstruction of *Sleeping Beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells. *Cell.* **91**, 501–510. doi: 10.1016/s0092-8674(00)80436-560
 60. Plasterk R.H., Izsvák Z., Ivics Z. (1999) Resident aliens: the *Tc1/mariner* superfamily of transposable elements. *Trends Genet.* **15**, 326–332. doi: 10.1016/s0168-9525(99)01777-1
 61. Arai Y., Hosoda F., Kobayashi H., Arai K., Hayashi Y., Kamada N., Kaneko Y., Ohki M. (1997) The inv(11)(p15q22) chromosome translocation of *de novo* and therapy-related myeloid malignancies results in fusion of the nucleoporin gene, *NUP98*, with the putative RNA helicase gene, *DDX10*. *Blood.* **89**, 3936–3944.
 62. Lee T.I., Young R.A. (2000) Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77–137. doi: 10.1146/annurev.genet.34.1.77
 63. Nigg E.A., Raff J.W. (2009) Centrioles, centrosomes, and cilia in health and disease. *Cell.* **139**, 663–678. doi: 10.1016/j.cell.2009.10.036
 64. Klug A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.* **79**, 213–231. doi: 10.1146/annurev-biochem-010909-095056
 65. Kumar M., Suleski J.E., Craig A.E., Kaspravic A.E., Sanderford M., Li M., Stecher G., Hedges S.B. (2022) TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* **39**, msac174. doi: 10.1093/molbev/msac174
 66. Cummings M.P. (1994) Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. *Trends Ecol. Evol.* **9**, 141–145. doi: 10.1016/0169-5347(94)90179-1
 67. Wallau G.L., Ortiz M.F., Loreto E.L. (2012) Horizontal transposon transfer in eukarya: detection, bias, and perspectives. *Genome Biol. Evol.* **4**, 689–699. doi: 10.1093/gbe/evs055
 68. Jangam D., Feschotte C., Betrán E. (2017) Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet.* **33**, 817–831. doi: 10.1016/j.tig.2017.07.011
 69. Hunter D.J., Williams K., Cartinhour S., Herrick G. (1989) Precise excision of telomere-bearing transposons during *Oxytricha fallax* macronuclear development. *Genes Dev.* **3**, 2101–2112. doi: 10.1101/gad.3.12b.210170
 70. Chen X., Landweber L.F. (2016) Phylogenomic analysis reveals genome-wide purifying selection on *TBE* transposons in the ciliate *Oxytricha*. *Mob. DNA.* **7**, 2. doi: 10.1186/s13100-016-0057-9
 71. Jahn C.L., Doktor S.Z., Frels J.S., Jaraczewski J.W., Krikau M.F. (1993) Structures of the *Euplotes crassus* *Tec1* and *Tec2* elements: identification of putative transposase coding regions. *Gene.* **133**, 71–78. doi: 10.1016/0378-1119(93)90226-s
 72. Doak T.G., Witherspoon D.J., Jahn C.L., Herrick G. (2003) Selection on the genes of *Euplotes crassus* *Tec1* and *Tec2* transposons: evolutionary appearance of a programmed frameshift in a *Tec2* gene encoding a tyrosine family site-specific recombinase. *Eukaryot. Cell.* **2**, 95–102. doi: 10.1128/EC.2.1.95-102.2003

Structure and Evolution of DNA Transposons of the *L31* Superfamily of Bivalves

M. V. Puzakov¹*, and L. V. Puzakova¹

¹ Kovalevsky Institute of Biology of the Southern Seas, Russian Academy of Sciences, Sevastopol, 299011 Russia

* e-mail: puzakov@ngs.ru

DNA transposons of the *IS630/Tc1/mariner* (*ITm*) are widespread representatives of DNA transposons that make a significant contribution to the evolution of eukaryotic genomes. With the start of large-scale application of next generation sequencing (NGS) technologies and the emergence of many new whole genome sequences of organisms in nucleotide collections, *ITm* elements have been identified in most taxa of the eukaryotic tree of life. Despite the rather detailed study of the diversity of *ITm* representatives, elements are still found that contribute to the expansion and revision of the classification of this group of DNA transposons. This paper presents for the first time a detailed analysis of the *L31* elements of bivalves, which resulted in a description of the structure, diversity, distribution, and phylogenetic position among the *ITm* elements. It was found that *L31* transposons are an independent superfamily in the *ITm* group, which has an ancient origin. Within the *L31* clade, rather high diversity was observed: five phylogenetic clusters were identified. At the moment, the presence of *L31* transposons in molluscs has been revealed only in bivalves in the subclass Autobranchia, with a predominance in diversity and quantity in the infraclass Pteriomorphia. It has also been shown that the protein encoded by the second open reading frame (ORF2) is an integral structural component of almost all full-length *L31* elements. The data obtained contribute to a better understanding of the evolution of representatives of *ITm* transposons. Further study of *L31* transposons in other taxa (cnidaria), as well as the study of the function of the second ORF protein, will provide an opportunity to better understand the evolution of DNA transposons, the mechanisms of horizontal transfer, and the contribution to eukaryotic biodiversity.

Keywords: DNA transposons, *L31* transposons, bivalves, genome evolution, biodiversity