

РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ

Федеральное государственное бюджетное учреждение науки
Институт цитологии и генетики

На правах рукописи

Сормачева Ирина Дмитриевна

**ЭВОЛЮЦИЯ И РАСПРОСТРАНЕНИЕ МОБИЛЬНЫХ
ГЕНЕТИЧЕСКИХ ЭЛЕМЕНТОВ В ГЕНОМАХ
ПРЕДСТАВИТЕЛЕЙ ОТРЯДА LEPIDOPTERA**

03.01.09 – математическая биология, биоинформатика

Диссертация на соискание ученой степени
кандидата биологических наук

Научный руководитель:
кандидат биологических наук
Блинов Александр Геннадьевич

Новосибирск – 2014

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	6
ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ.....	11
1.1. Понятие «мобильные генетические элементы»	11
1.2. Классификация мобильных генетических элементов эукариот	12
1.3. Класс I – Ретротранспозоны.....	16
1.3.1. Порядок non-LTR-ретротранспозоны	16
1.3.2. Общая структура non-LTR ретротранспозонов	17
1.3.3. Филогенетические взаимоотношения и классификация non-LTR ретротранспозонов	20
1.3.4. Разнообразие и распространение non-LTR ретротранспозонов.....	22
1.3.5. Перемещение non-LTR ретротранспозонов (ретротранспозиция).....	26
1.4. Класс II – ДНК транспозоны.....	30
1.4.1. Порядок TIR. Общая структура TIR ДНК транспозонов.....	31
1.4.2. Классификация TIR ДНК транспозонов.....	33
1.4.3. Разнообразие и распространение TIR ДНК транспозонов	35
1.4.4. Перемещение TIR ДНК транспозонов (транспозиция).....	36
1.5. Жизненный цикл мобильных генетических элементов	38
1.5.1. Снижение негативного воздействия МГЭ на геном хозяина	39
1.5.2. Горизонтальный перенос	40
1.5.2.1. Понятие «горизонтальный перенос»	40
1.5.2.2. Горизонтальный перенос МГЭ между эукариотическими организмами	41
1.5.2.3. Механизм горизонтального переноса.....	43
1.5.2.4. Влияние горизонтального переноса на МГЭ и геном хозяина	44
1.6. Влияние мобильных генетических элементов на геном хозяина	45
1.6.1. Воздействие МГЭ на гены генома хозяина	45
1.6.2. МГЭ и изменения размеров генома	46

1.6.3. МГЭ и хромосомные перестройки	46
1.6.4. Молекулярная доместикация генов МГЭ	47
1.7. Методы изучения мобильных генетических элементов.....	48
1.7.1. Методы поиска нуклеотидных и аминокислотных последовательностей МГЭ в базах данных и прочитанных геномах	49
1.7.2. Методы сравнительного анализа последовательностей МГЭ.....	51
1.7.3. Методы филогенетического анализа последовательностей МГЭ.....	51
1.7.4. Методы анализа горизонтального переноса МГЭ	53
Заключение по обзору литературы	54
ГЛАВА 2. МАТЕРИАЛЫ И МЕТОДЫ.....	56
2.1. Выбор стратегии исследования	56
2.2. Биоинформатический анализ	57
2.2.1. Поиск последовательностей <i>CR1</i> non-LTR ретротранспозонов и <i>mariner</i> -like ДНК транспозонов	57
2.2.2. Сравнительный анализ последовательностей МГЭ	58
2.2.3. Филогенетический анализ последовательностей МГЭ	58
2.2.4. Анализа горизонтального переноса МГЭ.....	59
2.3. Экспериментальные методики.....	60
2.3.1. Выделение геномной ДНК	60
2.3.2. Полимеразная цепная реакция.....	61
2.3.3. Электрофорез продуктов ПЦР в агарозном геле	61
2.3.4. Клонирование фрагментов ДНК.....	62
2.3.4.1. Лигирование фрагментов с Т-вектором	62
2.3.4.2. Трансформация компетентных клеток <i>E.coli</i>	62
2.3.5. Отбор и анализ клонов-трансформантов	62
2.3.6. Установление нуклеотидных последовательностей.....	63
2.3.7. Дот-блот гибридизация и создание библиотеки клонов	64
ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ	65

3.1. *CR1* non-LTR ретротранспозоны в геномах представителей отряда

Lepidoptera	65
3.1.1. Поиск <i>CR1</i> non-LTR ретротранспозонов в геномах представителей отряда Lepidoptera	66
3.1.2. Филогенетический анализ <i>CR1</i> non-LTR ретротранспозонов из геномов представителей отряда Lepidoptera	71
3.1.3. Распространение <i>CR1</i> элементов в геномах представителей отряда Lepidoptera	73
3.1.4. Вертикальная эволюция и горизонтальный перенос <i>CR1</i> элементов в геномах представителей отряда Lepidoptera	74
3.1.5. Поиск химерных <i>CR1B/MLE</i> элементов в геноме <i>B. mori</i>	77

3.2. *Mariner*-like TIR ДНК транспозоны в геномах представителей отряда

Lepidoptera	78
3.2.1. Поиск <i>MLE</i> ДНК транспозонов в геномах представителей класса Insecta ..	79
3.2.2. Поиск <i>MLE</i> ДНК транспозонов в геноме <i>B. mori</i>	82
3.2.3. Экспериментальный поиск <i>MLE</i> ДНК транспозонов в геномах представителей отряда Lepidoptera	84
3.2.4. Анализ <i>MLE</i> ДНК транспозонов в геномах представителей отряда Lepidoptera	84
3.2.5. Эволюция <i>MLE</i> ДНК транспозонов в геномах представителей отряда Lepidoptera	88
3.2.6. Экспериментальный поиск <i>Bmmar</i> -like элементов в геномах представителей отряда Lepidoptera	90
3.2.7. Неравномерное распространение <i>Bmmar</i> -like элементов в геномах представителей отряда Lepidoptera	93
3.2.8. Анализ <i>Bmmar</i> -like элементов в геномах представителей отряда Lepidoptera	95
3.2.9. Подтверждение горизонтального переноса <i>Bmmar</i> -like элементов в геномах представителей отряда Lepidoptera	98

3.3. Совместный горизонтальный перенос *CR1B* non-LTR ретротранспозонов и *mariner*-like ДНК транспозонов..... 101

3.3.1. Поиск химерных <i>CRIB/MLE</i> элементов в геномах представителей отряда <i>Lepidoptera</i>	102
3.3.2. <i>MLE</i> ДНК транспозоны в качестве векторов горизонтального переноса <i>CRIB non-LTR</i> ретротранспозонов.....	103
3.3.3. Механизм горизонтального переноса <i>CRIB non-LTR</i> ретротранспозонов и <i>MLE</i> ДНК транспозонов	106
ЗАКЛЮЧЕНИЕ	108
ВЫВОДЫ	110
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ	111
СПИСОК ЛИТЕРАТУРЫ	114
ПРИЛОЖЕНИЯ	133
Приложение 1	133
Приложение 2	135

ВВЕДЕНИЕ

Актуальность темы исследования

Мобильные генетические элементы (МГЭ) – последовательности ДНК, способные менять свою локализацию в геноме. МГЭ описаны практически для всех эукариотических организмов и могут составлять значительную фракцию их геномов.

МГЭ являются ключевыми факторами нестабильности геномов эукариот и участвуют в изменении структуры и размера геномов, нарушении функций генов и появлении новых функций. Активные исследования МГЭ показали, что с ними связано развитие системы поддержания продолжительности жизни клеток, формирование процесса V(D)J-рекомбинации, возникновение некоторых заболеваний человека, процесс старения, и многие другие свойства организмов-хозяев (Feschotte, Pritham, 2007; St. Laurent *et al.*, 2010; Rostant *et al.*, 2012; Kaer, Speek, 2013).

Благодаря способности к горизонтальному переносу, МГЭ могут участвовать в передаче генетической информации между репродуктивно изолированными и эволюционно удаленными видами. Особый интерес у исследователей горизонтального переноса вызывают МГЭ порядка non-LTR ретротранспозонов (класс I ретротранспозоны) и TIR ДНК транспозонов (класс II ДНК транспозоны). Горизонтальный перенос характерен для TIR ДНК транспозонов, на данные элементы приходится около 40% от всех случаев горизонтального переноса МГЭ, выявленных между геномами эукариот. Горизонтальный перенос non-LTR ретротранспозонов напротив чрезвычайно редкое явление, и наследование данной группы элементов происходит вертикально от родителей к потомкам. Изучение данных групп элементов позволит получить новые данные о разнообразии, особенностях распространения МГЭ и механизме горизонтального переноса.

Несмотря на актуальность проблемы изучения МГЭ, в картине происхождения, разнообразия и распространения МГЭ остается ряд нерешенных вопросов. Изучение данных вопросов осложняется тем, что в литературе накоплены данные о МГЭ, представленных в геномах модельных объектов, таких как *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Escherichia coli*,

Caenorhabditis elegans, *Arabidopsis thaliana* и *Homo sapiens*. Исследования особенностей МГЭ на новых «немодельных» объектах позволит получить дополнительную информацию об эволюции геномов хозяев и МГЭ в целом.

Отряд Lepidoptera (Чешуекрылые) - самый крупный отряд насекомых (Insecta), включающий в себя 47 суперсемейств, 124 семейства, 332 подсемейства и более 157000 видов (Kristensen, 2003; van Nieuwerkerken *et al.*, 2011). Представители данного отряда являются значимыми компонентами природных экосистем, оказывают влияние на деятельность человека и являются удобными объектами для научных исследований. Большинство представителей отряда имеют сложный жизненный цикл, что может представлять особый интерес в связи с изучением механизма горизонтального переноса МГЭ. Исследования МГЭ из геномов представителей отряд Lepidoptera позволят получить информацию о распространении, разнообразии и эволюции МГЭ элементов в геномах насекомых и эволюции геномов хозяев.

В рамках данной работы было проведено исследование распространения, разнообразия и эволюции *CRI* non-LTR ретротранспозонов и *mariner*-like TIR ДНК транспозонов в геномах представителей отряда Lepidoptera с помощью биоинформатических и экспериментальных методик. Применение экспериментальных подходов позволило получить последовательности МГЭ представителей отряда, для которых к настоящему моменту не секвенированы полногеномные последовательности. Биоинформатический анализ позволил осуществить поиск МГЭ в секвенированном геноме *Bombyx mori* и проанализировать последовательности, полученные в результате биоинформатического и экспериментального поиска.

Цель и задачи исследования

Целью данной работы являлось изучение распространения, разнообразия и эволюции *CRI* non-LTR ретротранспозонов и *mariner*-like TIR ДНК транспозонов в геномах представителей отряда Lepidoptera.

Для достижения данной цели были поставлены следующие задачи:

1. Исследование разнообразия *CR1* non-LTR ретротранспозонов в геноме тутового шелкопряда *Bombyx mori* и геномах других представителей отряда Lepidoptera (Insecta).
2. Анализ эволюции *CR1* non-LTR ретротранспозонов в геномах представителей отряда Lepidoptera.
3. Исследование разнообразия *mariner*-like TIR ДНК транспозонов в геноме тутового шелкопряда *Bombyx mori* и геномах других представителей отряда Lepidoptera и класса Insecta.
4. Анализ эволюции *mariner*-like TIR ДНК транспозонов в геномах представителей отряда Lepidoptera.
5. Выявление фактов возможного горизонтального переноса *mariner*-like TIR ДНК транспозонов и *CR1* non-LTR ретротранспозонов в геномах представителей отряда Lepidoptera.
6. Поиск и анализ химерных конструкций, содержащих *mariner*-like TIR ДНК транспозоны и *CR1* non-LTR ретротранспозоны, в геномах представителей отряда Lepidoptera.

Научная новизна

Впервые было проведено исследование распространения, разнообразия и эволюции *mariner*-like TIR ДНК транспозонов в геномах представителей отряда Lepidoptera. Был выявлен новый *mariner*-like ДНК транспозон *BmmarY* в геноме *Bombyx mori*. Подтвержден горизонтальный перенос *CR1B* non-LTR ретротранспозонов между бабочками семейства Lycaenidae и молями семейства Bombycidae. Продемонстрирован случай горизонтального переноса *mariner*-like ДНК транспозонов и выявлены химерные конструкторы, содержащие *mariner*-like ДНК транспозоны со встройкой *CR1B* элементов.

Теоретическая и практическая значимость

В результате проведенного исследования был обнаружен уникальный случай горизонтального переноса *mariner*-like ДНК транспозонов между представителями семейств Lycaenidae и Bombycidae. Был подтвержден горизонтальный перенос

CR1B non-LTR ретротранспозонов между представителями семейств *Lycaenidae* и *Bombycidae*. Были обнаружены свидетельства совместного горизонтально переноса *CR1B* и *mariner-like* элементов и участия *mariner-like* элементов в горизонтальном переносе *CR1B*.

Результаты, полученные в рамках настоящего исследования, вносят вклад в понимание распространения, разнообразия и эволюции *CR1B* non-LTR ретротранспозонов и *mariner-like* ДНК транспозонов в геномах представителей отряда *Lepidoptera*. Полученные данные могут быть внедрены в образовательный процесс и использованы при разработке векторных конструкций для трансгенеза, основанных на МГЭ насекомых.

Положение, выносимое на защиту

Распространение *Bmmar1-like*, *BmmarY-like* ДНК транспозонов и *CR1B* non-LTR ретротранспозонов в геномах представителей семейств *Lycaenidae* и *Bombycidae* произошло в результате горизонтального переноса между представителями данных семейств.

Апробация результатов

Материалы диссертации докладывались на Международной Московской конференции по компьютерной молекулярной биологии (Москва, Россия, 2011), на 6-ом Международном симпозиуме по Молекулярной Биологии Насекомых (Амстердам, Нидерланды, 2011), на 8-ой конференции Международной конференции по биоинформатике регуляции и структуры геномов и системной биологии (BGRS\SB'12, Новосибирск, 2012).

Личный вклад автора

Основные результаты работы получены автором самостоятельно. Экспериментальный поиск *CR1* non-LTR ретротранспозонов проводился совместно с к.б.н. Новиковой О.С. и Смышляевым Г.А. Биоинформатический анализ генома *Bombyx mori* осуществлялся в сотрудничестве с Новиковым А.С.

Публикации

По материалам диссертации было опубликовано 6 работ:

1. **Сормачева И.Д.**, Новиков А.С., Блинов А.Г. Распространение и филогенетические взаимоотношения *Tc1-Mariner* ДНК транспозонов в отряде *Lepidoptera* // Евразийский энтомологический журнал. 2010. Т. 9. № 3. С. 430-432.
2. **Sormacheva I.**, Novikov A., Blinov A. Vertical evolution and horizontal transmission of *Tc1/mariner* superfamily DNA transposons in Lepidopteran species // Proceedings of international Moscow conference on computational molecular biology - MCCMB'11. 2011. Moscow. Russia. P. 349-350.
3. **Sormacheva I.**, Novikov A., Blinov A. The evolution and diversity of *Tc1/mariner* superfamily DNA transposons in the genomes of Lepidoptera species // Proceedings of the Sixth International Symposium on Molecular Insect Science. 2011. Amsterdam, The Netherlands. P. 123.
4. **Сормачева И.Д.**, Блинов А.Г. LTR-ретротранспозоны растений. Вавиловский журнал Генетики и Селекции. 2011. Т. 15. №2. С. 351-381.
5. **Sormacheva I.**, Blinov A. The recurrent horizontal transfers of different transposable elements between Lepidoptera species // Proceedings of the eighth international conference on Bioinformatics of Genome Regulation and Structure\Systems Biology — BGRS\SB-2012. 2012. Novosibirsk, Russia. V. 1 P. 302.
6. **Sormacheva I.**, Smyshlyaev G., Mayorov V., Blinov A., Novikov A., Novikova O. Vertical Evolution and Horizontal Transfer of CR1 Non-LTR Retrotransposons and *Tc1/mariner* DNA Transposons in Lepidoptera Species // Mol. Biol. Evol. 2012. V. 29. P. 3685-3702.

Структура и объем работы

Диссертационная работа состоит из введения, обзора литературы, материалов и методов, результатов и обсуждений, выводов, списка сокращений и условных обозначений, списка литературы и приложений. Работа изложена на 137 страницах, содержит 30 рисунков и 3 таблицы. Библиографический указатель литературы включает 200 источников.

ГЛАВА 1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Понятие «мобильные генетические элементы»

Мобильные генетические элементы (МГЭ, синонимы: мобильные элементы, транспозоны, «прыгающие гены» - «jumping genes» и др.) – фрагменты ДНК, способные менять свою локализацию в геноме. В последовательностях МГЭ содержатся детальные инструкции, необходимые для перемещения элемента из одного сайта в другой сайт геномной ДНК клетки-хозяина.

МГЭ были открыты более полувека назад Барбарой МакКлинток в геноме кукурузы как генетические агенты, отвечающие за изменения пигментации зерен (McClintock, 1951). Позднее МГЭ были выявлены в геномах других модельных объектов, таких как, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Escherichia coli* и *Caenorhabditis elegans*. Активное развитие методов полногеномного секвенирования показало, что МГЭ являются значительными компонентами генетического материала большинства эукариотических организмов. Процентное содержание МГЭ в геномах представителей различных групп эукариотических организмов существенно варьирует. Так на мобильные генетические элементы приходится менее 3% генома рыбы *Fugu rubripes*, около 45% генома человека и от 50 до 90% геномов некоторых растений (Lander *et al.*, 2001; Meyers *et al.*, 2001; Aparicio *et al.*, 2002; Feschotte *et al.*, 2002; Schulman, Kalendar, 2005).

Длительное время МГЭ относили к категории так называемой «мусорной ДНК» («junk DNA»), функция которой была неизвестна. Однако, в настоящее время установлено, что МГЭ оказывают значительное влияние на эволюцию геномов хозяев. Перемещение МГЭ может приводить к нарушению функций генов, изменению регуляции генов, хромосомным перестройкам, приобретению новых функций, изменению размера геномов, обмену генетической информацией между репродуктивно изолированными видами и др. (Chenais *et al.*, 2012).

В настоящее время МГЭ не только активно изучаются, но и используются в качестве молекулярных маркеров в области филогении различных видов, генетических инструментов в области генной инженерии для введения

генетического материала и в качестве векторов при генотерапии заболеваний человека (Ohlfest *et al.*, 2004; Bowers *et al.*, 2006; Ott *et al.*, 2006; Aronovich *et al.*, 2007; Izsvak *et al.*, 2009; Peng *et al.*, 2009; Kalendar *et al.*, 2011; Di Matteo *et al.*, 2012).

1.2. Классификация мобильных генетических элементов эукариот

МГЭ значительно различаются по структуре и механизму перемещения, единственным общим свойством МГЭ является способность к перемещению. Первая классификация МГЭ эукариот была предложена Финнеганом и основывалась на различиях механизмов перемещения, в частности использования РНК или ДНК в качестве посредника (интермедиата) в процессе перемещения (Finnegan, 1989). На основе данного критерия выделяется два основных класса: класс I ретротранспозоны и класс II ДНК транспозоны (Рисунок 1.1).

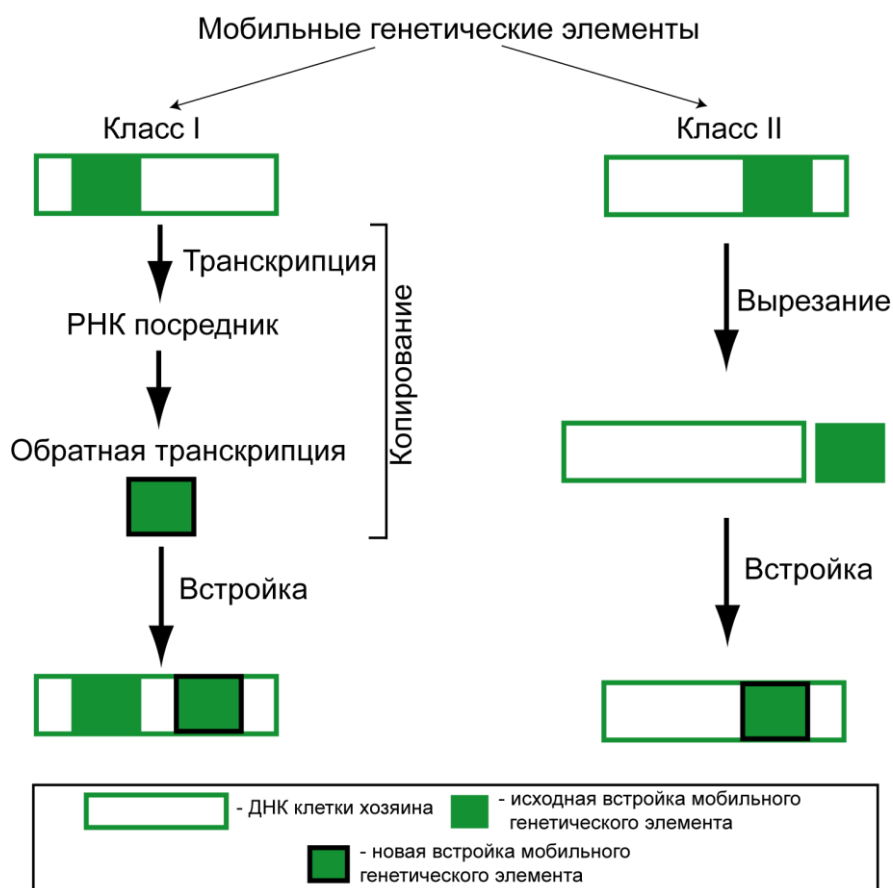


Рисунок 1.1. Классификация мобильных генетических элементов эукариот, основанная на особенностях механизма перемещения элементов (согласно Finnegan, 1989). В левой части рисунка представлена схема механизма «копирования-встройки», по которому происходит перемещение элементов класса I, в правой части представлена схема механизма «вырезания-встройки», по которому происходит перемещение элементов класса II.

Элементы класса I перемещаются по механизму «копирования – встройки» («copy and paste») и используют РНК в качестве посредника. В результате перемещения элементов класса I происходит увеличение количества копий ретротранспозона за счет образования новой копии мобильного элемента и сохранения исходной копии (репликативный механизм) (Рисунок 1.1). Элементы класса II вырезаются из одного сайта и встраиваются в новый сайт по механизму «вырезания – встройки» («cut and paste»), не используя РНК-посредник. Механизм «вырезания – встройки» является нерепликативным, таким образом, в отличие от элементов класса I в результате перемещения элементов класса II не происходит увеличения количества копий мобильного элемента (Рисунок 1.1). Обнаружение миниатюрных транспозирующих элементов, содержащих инвертированные повторы (miniature inverted repeat transposable elements – MITEs), перемещающихся по механизму «копирования – встройки», но не использующих РНК в качестве посредника, потребовало перехода к системе трех классов или классификации элементов в зависимости от ферментов, закодированных в их последовательности (Curcio, Derbyshire, 2003).

Иерархическая классификация, предложенная Вайкером и коллегами в 2007 году, учитывает оба подхода (Wicker *et al.*, 2007) (Рисунок 1.2). Данная классификация включает следующие таксономические единицы в иерархическом порядке: класс, подкласс, порядок, надсемейство, семейство и подсемейство.

На самом высоком уровне классификации («класс») МГЭ подразделяются в зависимости от использования/не использования РНК посредника в процессе перемещения (Рисунок 1.2). Таксономическая единица «подкласс» используется для разделения элементов класса II образующих и не образующих копию перед вырезанием из исходного (донорного) сайта. На уровне «порядков» МГЭ подразделяются в зависимости от особенностей механизма встройки, общей организации элемента и наличию последовательностей, кодирующих ферменты перемещения. Суперсемейства (надсемейства) в пределах одного порядка характеризуются сходной стратегией репликации, но отличаются по ряду характеристик.

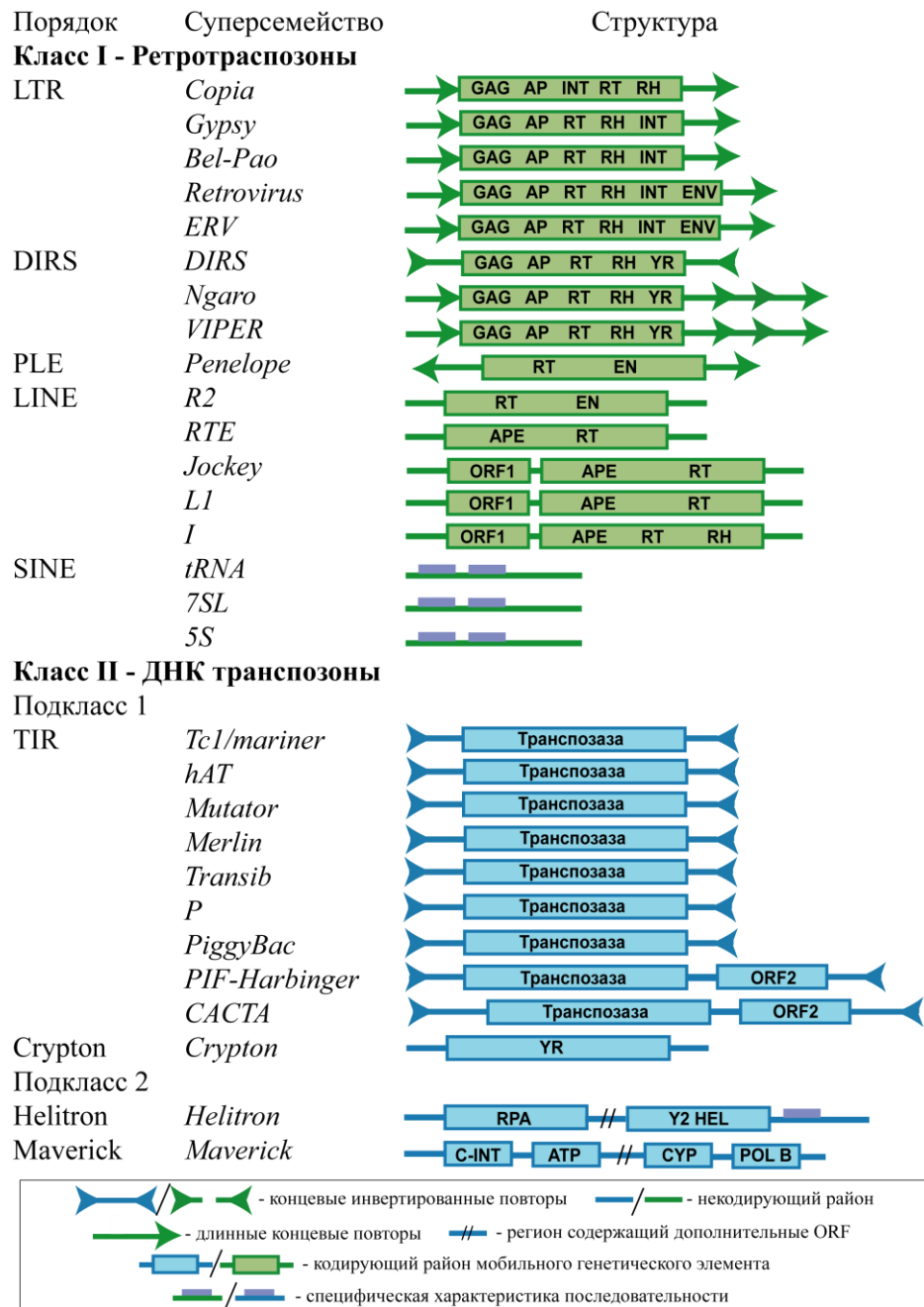


Рисунок 1.2. Современная универсальная классификация и структура мобильных генетических элементов эукариот (согласно Wicker *et al.*, 2007). Аббревиатура: AP (APE) – apurinic/apyrimidin endonuclease (апуриновая/апиримидиновая эндонуклеаза); ATP – ATPase (АТФаза); C-INT – C-integrase (интеграза типа c-int); CYP – Cysteine protease (цистеиновая протеаза); DIRS – Dictyostelium intermediate repeat sequence (переходные повторяющиеся последовательности *Dictyostelium*); EN – endonuclease domain (домен эндонуклеазы); ENV – Envelope protein (белок оболочки); HEL – DNA helicase (ДНК геликаза); HYR – tyrosine recombinase (тирозин рекомбиназа); INT – integrase (интеграза); LINE – long interspersed nuclear element (длинный диспергированный ядерный повтор); LTR – long terminal repeat (длинный концевой повтор); ORF – open reading frame (открытая рамка считывания); PLE – Penelope-like elements (Penelope – подобные элементы); RH – RNase H (RNH) (рибонуклеаза); RPA – replication protein A (белок инициации репликации); POL B – DNA polymerase B (ДНК полимеразы B), RT – reverse transcriptase domain (домен обратной транскриптазы); SINE – short interspersed nuclear element (короткий диспергированный ядерный повтор); TIR – terminal inverted repeat (концевые инвертированные повторы); TSD – target site duplication (дупликация сайта-мишени).

К таким характеристикам относятся структура кодирующих или некодирующих областей, наличие и размер дубликации сайта мишени (target site duplication – TSD). TSD представляет собой короткий прямой повтор, генерируемый при встройке справа и слева от последовательности элемента. Суперсемейства подразделяются на семейства согласно гомологии последовательностей элементов. Подсемейства определяются на основании результатов филогенетического анализа (Wicker *et al.*, 2007).

Автономные и не автономные элементы. Вне зависимости от особенностей механизма перемещения МГЭ можно классифицировать по способности к самостоятельному перемещению. Среди ретротранспозонов и ДНК транспозонов встречаются автономные и неавтономные элементы.

Автономные элементы способны к самостоятельному перемещению, в то время как неавтономные используют ферментную систему автономных элементов для перемещения (Feschotte *et al.*, 2002). Неавтономные элементы, как правило, представляют собой копии автономных элементов, накопившие различные мутации, но иногда они имеют лишь незначительное сходство по последовательности с их автономными предшественниками. Количество неавтономных элементов в геноме может значительно превышать количество автономных. Так, семейство неавтономных LTR ретротранспозонов *Dasheng* является наиболее высококопийным семейством МГЭ в геноме риса (Jiang *et al.*, 2002).

Активные и неактивные элементы. Активные элементы в отличие от неактивных способны перемещаться по геному и создавать собственные копии в новых сайтах мишенях. Как правило, активными являются полные автономные элементы, однако неавтономные копии элементов могут использовать ферментативную систему автономных и осуществлять активное перемещение в геноме. Так, неавтономные укороченные ДНК транспозоны группы MITE, сходство которых с ближайшими автономными элементами ограничивается последовательностью концевых инвертированных повторов (TIR – terminal inverted repeat), могут использовать ферменты автономных элементов для перемещения и увеличения числа копий в геноме (Feschotte *et al.*, 2003). Наиболее детально

изученной является система *mPing/Pong* из генома риса. Транспозиция MITE элемента *mPing* зависит от активности ДНК транспозона *Pong* суперсемейства *PIF/Harbinger* (Jiang *et al.*, 2003).

1.3. Класс I – Ретротранспозоны

Ретротранспозоны (класс I) – МГЭ, использующие РНК-посредник в процессе перемещения (ретротранспозиции). Путем обратной транскрипции с РНК-посредника синтезируется кДНК копия, которая впоследствии интегрируется в новый сайт генома. За полный цикл репликации образуется одна новая копия ретротранспозона. На основании различий структурной организации и филогении последовательностей ключевого фермента ретротранспозиции, обратной транскриптазы (reverse transcriptase – RT), в классе ретротранспозонов выделяется один подкласс и пять порядков: LTR ретротранспозоны, *DIRS*-подобные элементы, *Penelope*-подобные элементы (PLEs), LINE (non-LTR retrotransposons – non-LTR ретротранспозоны) и SINE элементы (short interspersed nuclear element - короткие диспергированные ядерные повторы) (Рисунок 1.2).

Наибольший интерес для исследователей МГЭ представляет порядок non-LTR-ретротранспозонов. Элементы данного порядка широко представлены в геномах позвоночных и беспозвоночных животных. В отличие от других представителей класса ретротранспозонов для non-LTR-ретротранспозонов ранее было описано исключительно вертикальное наследование: передача элементов от родителей к потомкам. Однако, для элементов данного порядка, был выявлен ряд исключений характеризующихся передачей элементов «горизонтально» между геномами репродуктивно изолированных видов (Zupunski *et al.*, 2001; Novikova *et al.*, 2007). Механизм данного явления до настоящего момента не установлен. Таким образом, изучение разнообразия и особенностей распространения non-LTR-ретротранспозонов является актуальной задачей.

1.3.1. Порядок non-LTR-ретротранспозоны

Non-LTR-ретротранспозоны или LINE элементы (Long Interspersed Nuclear Elements – длинные диспергированные ядерные элементы) – МГЭ, относящиеся к классу ретротранспозонов и не содержащие длинных концевых повторов,

характерных для других ретротранспозонов. Non-LTR ретротранспозоны широко распространены в геномах различных эукариотических организмов, но преобладают в геномах животных. Копийность и разнообразие non-LTR ретротранспозоны значительно варьируют в различных группах живых организмов. Элементы семейства L1 представлены большим количеством копий ($\sim 10^5$) в геномах млекопитающих и составляют 20% генома человека (Wicker *et al.*, 2007). С другой стороны в геноме москита *Anopheles gambiae* представлено около 100 различных семейств non-LTR ретротранспозонов, составляющих не более 3% генома (Biedler, Tu, 2003). В геномах растений non-LTR ретротранспозоны встречаются реже, чем LTR ретротранспозоны. Большинство изученных non-LTR-ретротранспозонов в геномах растений относятся к суперсемействам L1 и RTE (Zupunski *et al.*, 2001).

1.3.2. Общая структура non-LTR ретротранспозонов

Структурная организация трех модельных элементов порядка non-LTR ретротранспозонов (R2 из генома *Bombyx mori*, L1 из генома человека и I factor из генома *Drosophila melanogaster*) представлена на рисунке 1.3.

Non-LTR ретротранспозоны имеют размер от 3 до 8 тыс. п.о. Полноразмерные копии автономных non-LTR ретротранспозонов содержат следующие основные структурные компоненты: 5'-нетранслируемая область (5' UTR), центральная часть и 3'-нетранслируемая область (3' UTR). 5'- и 3'-нетранслируемые области элементов являются переменными участками (Рисунок 1.3).

5'-нетранслируемая область. Большинство 5' UTR содержат внутренний промотор РНК полимеразы II, однако встречаются элементы, не содержащие эндогенные промоторы. Так элементы R2, обнаруженные в локусах рибосомных генов насекомых, не содержат собственных промоторов и транскрибируются вместе с геном 28S рРНК (George, Eickbush, 1999) (Рисунок 1.3).

Центральная часть. В зависимости от структуры центральной части элемента выделяют две группы non-LTR ретротранспозонов: REL-endo элементы и ARE-элементы (Malik *et al.*, 1999) (Рисунок 1.3).

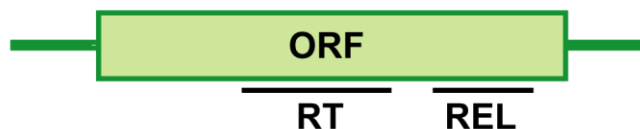
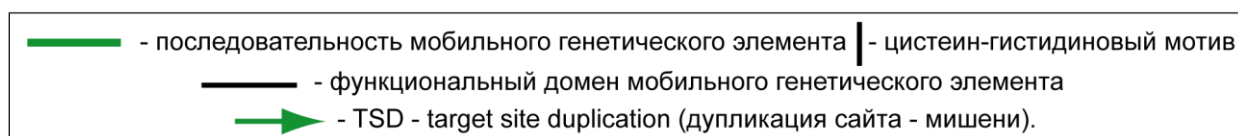
R2 (*Bombyx mori*)L1 (*Homo sapiens*)I factor (*Drosophila melanogaster*)

Рисунок 1.3. Структурная организация non-LTR ретротранспозонов (согласно Han, 2010) на примере типичных представителей данного порядка: R2 из генома *Bombyx mori*, L1 из генома человека и I factor из генома *Drosophila melanogaster*. Цветными прямоугольниками выделены открытые рамки считывания (ORF – open reading frame) non-LTR ретротранспозонов. Аббревиатура: APE – apurinic/apyrimidin endonuclease (апуриновая/апиримидиновая эндонуклеаза); BD – C-terminal nucleic acid binding domain (С – концевой домен связывающий нуклеиновые кислоты); LZ – leucine zipper domain (домен лейцинового zipper («молнии»)), RT – reverse transcriptase domain (домен обратной транскриптазы); REL – REL-endo restriction enzyme-like endonuclease domain (домен рестрикционной эндонуклеазы). $(A)_n$ – полиаденильный хвост, $(TAA)_n$ – тринуклеотидный повтор TAA.

REL-endo non-LTR ретротранспозоны или RE-type ретротранспозоны содержат одну открытую рамку считывания (ORF2), кодирующую два фермента – обратную транскриптазу и эндонуклеазу REL-endo (restriction enzyme-like endonuclease) (Malik *et al.*, 1999). Примером REL-endo non-LTR ретротранспозона является элемент R2 из генома *Bombyx mori* (Рисунок 1.3). Обратная транскриптаза представляет собой РНК-зависимую ДНК полимеразу. RT является ключевым ферментом обратной транскрипции и имеет общее происхождение для всех ретротранспозонов. Эндонуклеаза (EN) REL-endo ретротранспозонов наиболее близка к эндонуклеазе рестрикции типа IIS, содержащей отдельные ДНК-расщепляющий и ДНК-связывающий домены (Yang *et al.*, 1999). Наличие

рестрикционной эндонуклеазы обеспечивает сайт-специфическую встройку новых копий данных элементов (Eickbush, 2002). Основными сайтами-мишенями для интеграции REL-endo non-LTR ретротранспозонов являются микросателлитные и субтеломерные повторы, а также высокоповторенные гены, такие как гены рибосомной РНК (Xiong, Eickbush, 1993; Burke *et al.*, 1995; Eickbush, 2002).

Центральная часть **APE non-LTR ретротранспозонов** содержит две открытые рамки считывания (ORF1 и ORF2) (Moran, Gilbert, 2002). Примерами APE non-LTR ретротранспозонов являются элемент L1 из генома *Homo sapiens* и элемент *I* из генома *Drosophila melanogaster* (Рисунок 1.3). Первая рамка считывания, соответствующая гену *gag*, содержит мотивы, определяющие РНК-связывающую активность и шаперонную активность (Dawson *et al.*, 1997; Hohjoh, Singer, 1997; Kolosha, Martin, 2003). Для многих групп элементов в ORF1 были выявлены мотивы типа «цинковые пальцы» («Zn-finger»), обеспечивающие РНК-связывающую активность (Martin, Bushman, 2001).

Вторая открытая рамка считывания (ORF2) кодирует, по крайней мере, два фермента: обратную транскриптазу и эндонуклеазу (Dlakic, 2000; Hofmann *et al.*, 2000). Эндонуклеаза, кодируемая данной группой ретротранспозонов, наиболее близка к апуриновой/апиримидиновой эндонуклеазе (apurinic/apurimidin endonuclease - APE) и помимо эндонуклеазной активности обладает способностью к репарации ДНК (Martin *et al.*, 1995; Feng *et al.*, 1996; Weichenrieder *et al.*, 2004). Кроме перечисленных доменов в ORF2 ряда элементов закодированы домены, функции которых к настоящему моменту не установлены. Некоторые семейства элементов содержат С-концевой домен рибонуклеазы H (RNase H) (Eickbush, Jamburuthugoda, 2008).

3'-нетранслируемая область. 3'-нетранслируемая область non-LTR ретротранспозонов варьирует по длине и нуклеотидной последовательности. Как правило, данная область содержит специфические последовательности, узнаваемые обратной транскриптазой (Kajikawa, Okada, 2002; Osanai *et al.*, 2004; Anzai *et al.*, 2005). 3'-концевая часть non-LTR ретротранспозов может содержать последовательность богатую повторами аденозина (полиА (polyA)) или сигнал

полиаденилирования (Blinov *et al.*, 1997; Chambeyron *et al.*, 2002; Kajikawa, Okada, 2002).

Индивидуальные non-LTR ретротранспозоны могут обладать структурными особенностями, отличными от вышеописанных. Так, в ORF1 ряда элементов группы APE non-LTR ретротранспозонов закодирован домен эстеразы, который, по-видимому, может участвовать в процессе ретротранспозиции (Kapitonov, Jurka, 2003). Кроме того, большинство копий non-LTR ретротранспозонов содержат делеции в области 5'-UTR, таким образом, лишь часть non-LTR ретротранспозонов являются полноразмерными и активными (Szak *et al.*, 2002). Например, из 500000 копий элементов L1 в геноме человека только около 7000 являются полноразмерными, и около 90 из них являются активными (Brouha *et al.*, 2003).

1.3.3. Филогенетические взаимоотношения и классификация non-LTR ретротранспозонов

Единственным общим элементом структуры всех non-LTR ретротранспозонов является домен обратной транскриптазы. Высокая консервативность семи каталитических мотивов обратной транскриптазы, позволила провести сравнения данных последовательностей между различными non-LTR ретротранспозонами. В первоначальных исследованиях с помощью филогенетического анализа последовательностей обратной транскриптазы в группе non-LTR ретротранспозонов был выделено 11 групп элементов (Malik *et al.*, 1999). Активные исследования разнообразия non-LTR ретротранспозонов в геномах различных организмов позволили выявить новые элементы и в настоящее время в группе non-LTR ретротранспозонов выделяют 5 кластеров и 23 филогенетические группы (Novikova, Blinov, 2009) (Рисунок 1.4).

По результатам филогенетического анализа были выявлены принципиальные этапы эволюции non-LTR ретротранспозонов: приобретение дополнительной открытой рамки считывания (ORF1) и новых функциональных доменов (Malik *et al.*, 1999) (Рисунок 1.4). Эволюционно более древние non-LTR ретротранспозоны REL-endo (GENIE, CRE, R2, NeSL, HERO и R4) содержат одну открытую рамку (ORF2) считывания и кодируют домен обратной транскриптазы и эндонуклеазы

ретрикции (REL-endo). Древние non-LTR ретротранспозоны, по-видимому, произошли от интронов группы II бактерий (Martínez-Abarca, Toro, 2000) или симбиотических органелл и благодаря наличию рестрикционной эндонуклеазы обладали возможностью специфического встраивания в геном.

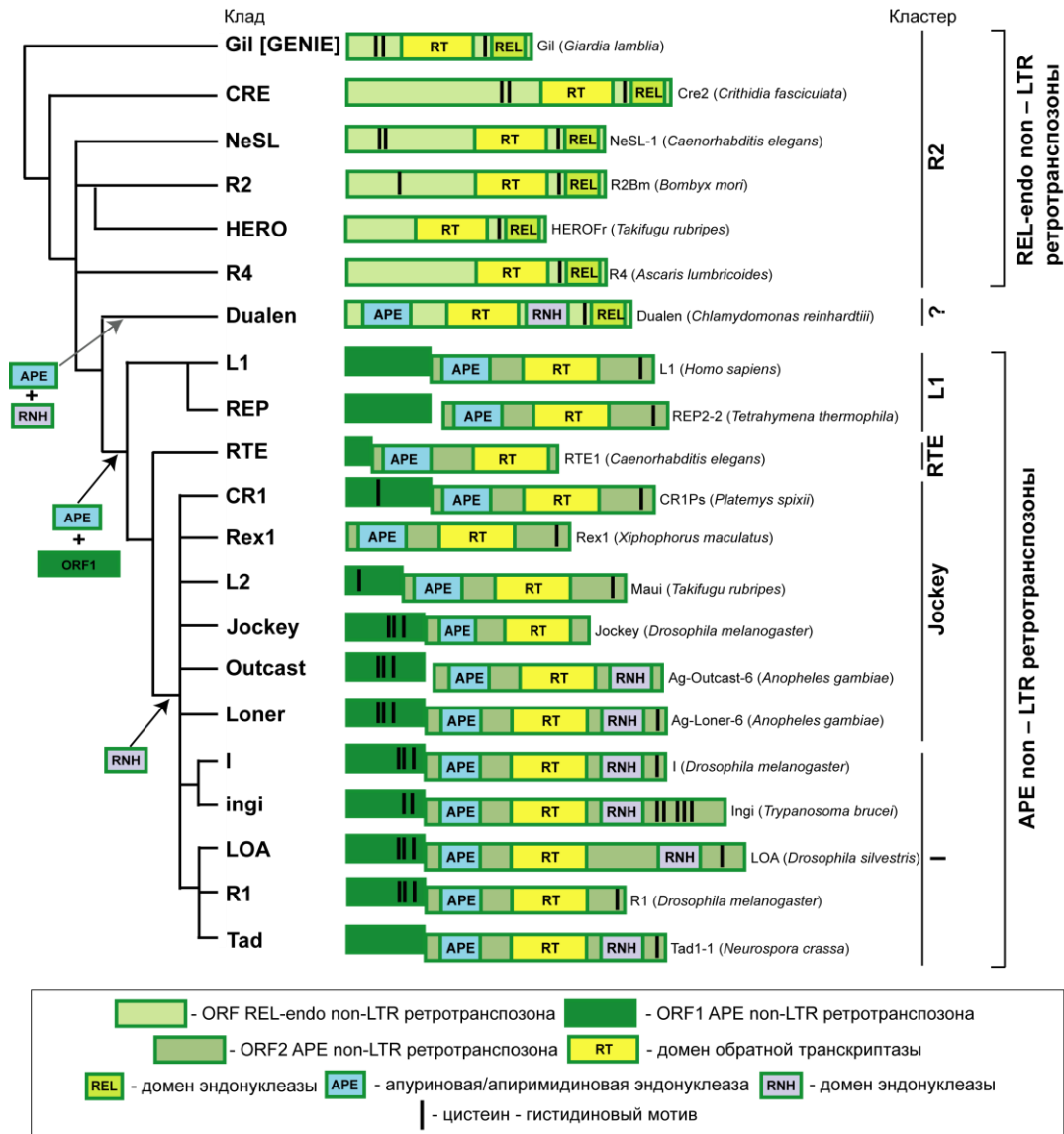


Рисунок 1.4. Филогенетическое древо и структурная организация типичных представителей различных клад non-LTR ретротранспозонов (согласно Malik *et al.*, 1999 и Novikova, Vlinov, 2009). В левой части рисунка изображена филогения клад non-LTR ретротранспозонов. В правой части рисунка изображены структуры non-LTR ретротранспозонов соответствующих клад. ORF1 и ORF2 элементов клад Outcast и REP изображены со смещением поскольку находятся в разных рамках. Стрелками указано предполагаемое время появления дополнительных доменов в структуре non-LTR ретротранспозонов. Аббревиатура: APE – apurinic/apurimidin endonuclease (домен апуриновой/апиримидиновой эндонуклеазы); ORF – open reading frame (открытая рамка считывания); REL – REL-endo restriction enzyme-like endonuclease domain (домен рестрикционной эндонуклеазы); RNH – домен рибонуклеазы H; RT – reverse transcriptase domain (домен обратной транскриптазы).

В дальнейшем REL эндонуклеаза была заменена на APE эндонуклеазу, приобретенную из генома хозяина. Данная замена произошла на ранней стадии эволюции non-LTR ретротранспозонов, на уровне предшественников суперсемейства L1 (Рисунок 1.4). Утратив эндонуклеазный домен REL, ретротранспозоны утратили способность к сайт-специфической встройке в геном. Одновременно с приобретением APE домена была приобретена вторая открытая рамка считывания (ORF1), которая сохранилась у элементов всех кластеров APE non-LTR ретротранспозонов, кроме RTE (Malik, Eickbush, 1998) (Рисунок 1.4).

ORF1 позволила non-LTR ретротранспозонам приобрести дополнительные функции и является обязательной для ретротранспозиции APE-элементов. Домен РНКазы H, представленный у элементов филогенетических групп Loner, Outcast, Tad, ingi, LOA, R1 и I, был заимствован из генома хозяина через некоторое время после приобретения эндонуклеазы APE (Malik *et al.*, 1999). В группах R1 и Tad присутствуют как элементы, содержащие домен РНКазы H, так и элементы, не содержащие данного домена.

1.3.4. Разнообразие и распространение non-LTR ретротранспозонов

Древние REL-endo non-LTR ретротранспозоны. К эволюционно более древним non-LTR ретротранспозонам относятся элементы шести филогенетических групп: GENIE, CRE, R2, NeSL, HERO и R4 (Malik *et al.*, 1999; Eickbush, Malik, 2002; Novikova, Vlinov, 2009) (Рисунок 1.5).

Элементы филогенетических групп Gil [GENIE] и CRE обнаружены в геномах представителей царства Простейшие (Protozoa). Филогенетическая группа Gil [GENIE] считается одной из наиболее древних групп non-LTR ретротранспозонов, представители данной филогенетической группы были выявлены исключительно в геноме примитивного простейшего *Giardia lamblia* (Arkhipova, Morrison, 2001) (Рисунок 1.5).

Представители филогенетической группы CRE выявлены в геномах паразитических простейших, таких как *Trypanosoma brucei*, *Trypanosoma cruzi*, *Crithidia fasciculata* и *Leptomonas seymouri* и гриба *Cryptococcus neoformans* (отдел Basidiomycota) (Bellofatto *et al.*, 1988; Aksoy *et al.*, 1990; Villanueva *et al.*, 1991; Goodwin, Poulter, 2001) (Рисунок 1.5).

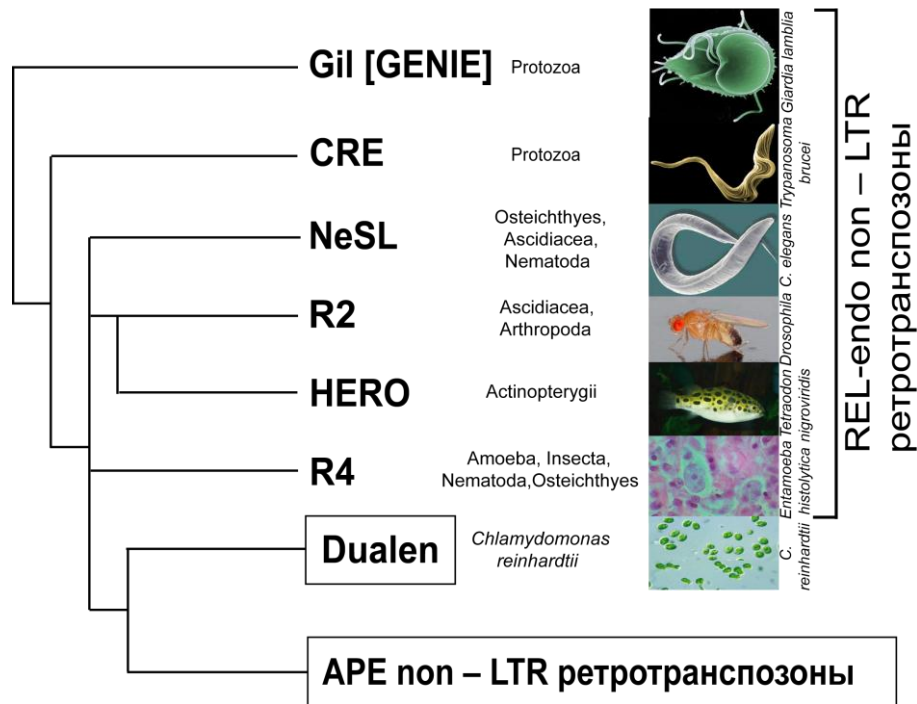


Рисунок 1.5. Филогенетическое древо и распространение REL-endo non-LTR ретротранспозонов. В левой части рисунка изображена филогения клад REL-endo non-LTR ретротранспозонов. В правой части рисунка приведены группы организмов, в геномах которых распространены REL-endo non-LTR ретротранспозонов.

Представители группы NeSL выявлены в геномах беспозвоночных (нематод, Nematoda), хордовых (асцидий, класс Ascidiacea) и костных рыб (класс Osteichthyes) (Eickbush, Malik, 2002) (Рисунок 1.5).

Впервые элементы клада R2 были обнаружены у *Drosophila*, затем они были выявлены в геномах асцидий *Ciona intestinalis* и многих членистоногих (тип Arthropoda) (Eickbush *et al.*, 1997; Permanyer *et al.*, 2003). Элементы клада HERO широко представлены в геномах рыб (надкласс Pisces, класс Actinopterygii) и, по-видимому, являются одним из семейств клада R2 (Kojima, Fujiwara, 2005).

Элементы клада R4 описаны для широкого круга эукариотических организмов: амёб (Sharma *et al.*, 2001), насекомых (класс Insecta) (Xiong, Eickbush, 1993; Kojima, Fujiwara, 2004); нематод (Burke *et al.*, 1995), костных рыб (Volf *et al.*, 2001) (Рисунок 1.5).

Молодые APE non-LTR ретротранспозоны. К APE ретротранспозонам относятся четыре кластера и шестнадцать филогенетических групп (Malik *et al.*, 1999; Eickbush, Malik, 2002; Zingler *et al.*, 2005, Novikova, Blinov, 2009) (Рисунок 1.6).

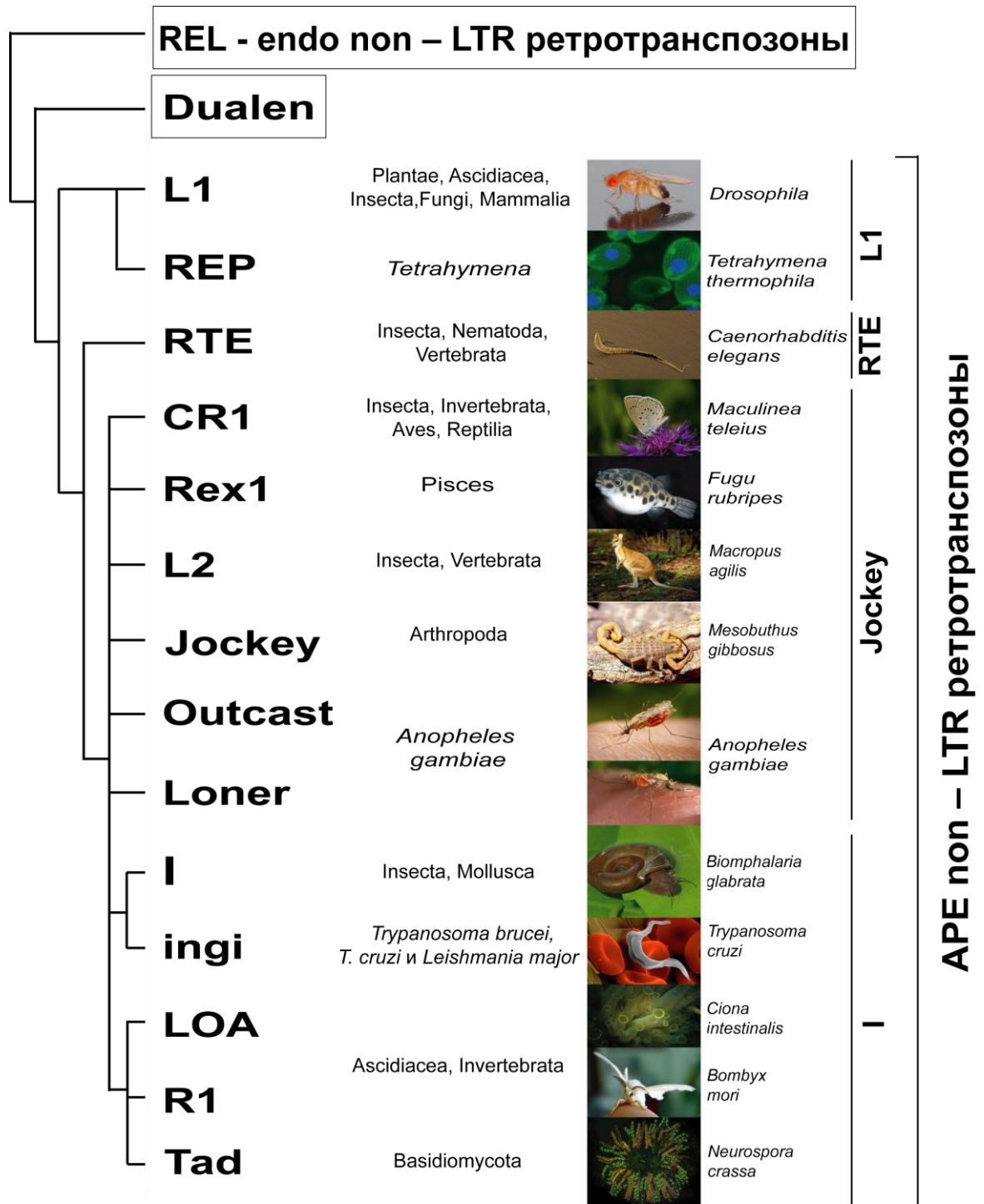


Рисунок 1.6. Филогенетическое древо и распространение APE non-LTR ретротранспозонов. В левой части рисунка изображена филогения клада APE non-LTR ретротранспозонов. В правой части рисунка приведены группы организмов, в геномах которых распространены APE non-LTR ретротранспозонов.

L1 кластер. Кластер L1 считается наиболее древним среди APE non-LTR ретротранспозонов и широко распространен в различных группах эукариотических организмов. Несмотря на наличие APE эндонуклеазы, ряд элементов данного кластера сохранил способность к специфической интеграции в геном хозяина.

Элементы семейства *Tall* данного кластера являются основными представителями non-LTR ретротранспозонов выявленными в геномах растений (Malik *et al.*, 1999). Кроме того, элементы L1 представлены в геномах насекомых (Berezikov *et al.*, 2000), грибов (Goodwin *et al.*, 2001), асцидий (Permanyer *et al.*, 2003), приматов, человека и других млекопитающих (Babushok, Kazazian, 2007) (Рисунок 1.6). При изучении свободноживущих эукариотических микроорганизмов *Tetrahymena thermophila* был выявлена новая филогенетическая группа REP кластера L1 (Fillingham *et al.*, 2004). В настоящее время статус данной группы уточняется (Рисунок 1.6).

RTE кластер. Кластер RTE образован единственной филогенетической группой, элементы которой содержат одну рамку считывания и характеризуются значительными делециями в 5' области (Malik, Eickbush, 1998; Zingler *et al.*, 2005) (Рисунок 1.4). RTE элементы присутствуют в геномах насекомых, беспозвоночных (нематод) и позвоночных (Malik, Eickbush, 1998; Zupunski *et al.*, 2001; Zingler *et al.*, 2005) (Рисунок 1.6).

Jockey кластер. Jockey кластер содержит элементы следующих филогенетических групп: Jockey, CR1, Rex1, L2, Loner и Outcast. Элементы клада Jockey были обнаружены исключительно в геномах членистоногих (Zingler *et al.*, 2005). CR1 элементы выявлены в геномах насекомых, беспозвоночных, птиц и рептилий (Burch *et al.*, 1993; Marin *et al.*, 1998; Zingler *et al.*, 2005; Shedlock, 2006). Rex1 является группой non-LTR ретротранспозонов, специфичной для рыб (Volff *et al.*, 2000). L2 элементы характерны для различных групп насекомых и позвоночных (Lovsin *et al.*, 2001). Элементы филогенетических групп Loner и Outcast обнаружены исключительно у представителей *Anopheles gambiae* (Biedler, Tu, 2003).

I кластер. В кластере I выделяют пять крупных филогенетических групп: Tad1, R1, LOA, I и ingi. Элементы клада Tad описаны только в геномах грибов (Malik *et al.*, 1999). Элементы клада R1, LOA, I изначально считались специфическими группами ретротранспозонов беспозвоночных, однако впоследствии были выявлены в геноме асцидии *C. intestinalis* (Malik *et al.*, 1999; Permanyer *et al.*, 2003). I элементы представлены в геномах различных насекомых и

в геноме моллюска *Biomphalaria glabrata* (Mollusca). Представители группы *ingi* выявлены исключительно в геномах паразитических простейших *Trypanosoma brucei*, *Trypanosoma cruzi* и *Leishmania major* (Aksoy *et al.*, 1990; Villanueva *et al.*, 1991; Bringaud *et al.*, 2006). Элементы *ingi* составляют значительную фракцию геномов данных организмов (до 6%).

Deceiver и Inkcap. Два новых клада Deceiver и Inkcap APE non-LTR ретротранспозонов были выявлены в геномах грибов группы Basidiomycota (Novikova *et al.*, 2009). Данные группы отсутствуют на рисунке 1.6, так как их таксономический статус к настоящему моменту не установлен.

Группа Dualen. В геноме зеленой водоросли *Chlamydomonas reinhardtii* были выявлены уникальные элементы группы Dualen (Kojima, Fujiwara, 2005) (Рисунок 1.5). Данные элементы содержат в своей структуре одну ORF, в которой закодированы домены обратной транскриптазы, REL-endo и APE эндонуклеаз и РНКазы Н (Рисунок 1.4). Таким образом, группа Dualen представляет собой промежуточную (переходную) форму между REL-endo и APE non-LTR ретротранспозонами. Кроме того, наличие домена РНКазы Н свидетельствует о том, что данный домен был приобретен non-LTR ретротранспозонами раньше, чем образовался кластер I (Malik *et al.*, 1999; Kojima, Fujiwara, 2005). Альтернативным объяснением данного факта является возможность неоднократного приобретения домена РНКазы Н в эволюции non-LTR ретротранспозонов.

1.3.5. Перемещение non-LTR ретротранспозонов (ретротранспозиция)

Полный цикл ретротранспозиции non-LTR ретротранспозонов включает следующие стадии: транскрипция, трансляция, сборка рибонуклеопротеидных комплексов, транспорт в ядро клетки-хозяина, обратная транскрипция, интеграция в геном.

Экспрессия non-LTR ретротранспозонов. Для транскрипции и трансляции non-LTR ретротранспозоны используют ферментативный аппарат клетки-хозяина. Транскрипция non-LTR ретротранспозонов осуществляется РНК-полимеразой II или РНК-полимеразой I с внутреннего промотора, или промотора соседнего хозяйского гена (Han, 2010). Для транскриптов, полученных с помощью РНК-

полимеразы II, инициация трансляции ORF1, по видимому, осуществляется по стандартному «сканирующему» механизму инициации. Подтверждения для данного предположения были получены в экспериментах по изучению экспрессии L1 элемента (Han, Воеке, 2004). Инициация трансляции элементов, транскрибируемых с помощью РНК полимеразы I, таких как R1/R2, может осуществляться за счет последовательностей IRES (Internal Ribosomal Entry Site - участок внутренней посадки рибосомы) (George, Eickbush, 1999).

В случае элементов содержащих две рамки считывания, трансляция ORF2 может осуществляться путем реинициации синтеза белка или сдвига рамки считывания. Первый механизм предполагает независимую трансляцию белков ORF1p и ORF2p, второй напротив, обеспечивает синтез объединенного полипротеина ORF1p - ORF2p.

Сборка рибонуклеопротеидных комплексов. Белки и транскрипты non-LTR ретротранспозонов собираются в рибонуклеопротеидные комплексы (ribonucleoprotein particles-RNPs), локализованные преимущественно в цитоплазме клетки-хозяина. В исключительных случаях данные комплексы могут быть локализованы в ядре: RNPs элемента L1 клеточных линий хронического миелоидного лейкоза крыс, RNPs элементов интегрирующихся в область теломерных повторов (Matsumoto *et al.*, 2004; Kirilyuk *et al.*, 2008).

Транспорт рибонуклеопротеидных комплексов. Последующий транспорт RNPs комплексов в ядро, по-видимому, опосредован цитоскелетом клетки-хозяина. В исследованиях транспорта RNPs элемента L1 было показано взаимодействие ORF1p в составе RNPs с филаментами цитоскелета клетки (Kirilyuk *et al.*, 2008). Паттерн переноса в ядро RNPs I элемента у *Drosophila* совпадает с транспортом RNPs белка *bicoid* (Seleme *et al.*, 1999). Вышеперечисленные факты подтверждают возможность транспорта RNPs non-LTR ретротранспозонов из цитоплазмы в ядро за счет элементов цитоскелета клетки-хозяина.

Обратная транскрипция. Согласно современным представлениям, обратная транскрипция non-LTR ретротранспозонов происходит по механизму TPRT – target-primed reverse transcription (Han, 2010). TPRT модель обратной транскрипции была экспериментально подтверждена для R2 элемента тутового

шелкопряда *Bombyx mori* (Christensen, Eickbush, 2005). *In vivo* и *in vitro* исследования показали, что механизм обратной транскрипции других non-LTR ретротранспозонов согласуется с TPRT моделью (Chambeyron *et al.*, 2002; Cost *et al.*, 2002).

Модель TPRT реакции non-LTR ретротранспозонов. Механизм обратной транскрипции детально изучен для R2 REL-endo non-LTR ретротранспозона. В единственной рамке считывания данного элемента закодирован полипротеин, содержащий N-концевой ДНК связывающий домен, центральный домен обратной транскриптазы и С-концевой эндонуклеазный домен (Рисунок 1.7). Предполагается, что С-концевой домен содержит отдельные ДНК-связывающий и эндонуклеазный подмодули.

Полный комплекс ферментов R2 функционирует в виде димера и имеет две конформации, выполняющие различные каталитические функции. Комплекс ферментов R2 в конформации первого типа (первая субъединица) связывается с РНК транскриптом R2, а комплекс ферментов R2 в конформации второго типа (вторая субъединица) взаимодействует с первой субъединицей в области сайта мишени ДНК (Рисунок 1.7).

TPRT реакция состоит из 4 этапов (Christensen, Eickbush, 2005):

(1) Первая субъединица связывается с РНК транскриптом и геномной ДНК выше сайта мишени, и разрезает одну из цепей ДНК сайта мишени. Образовавшийся свободный 3'-ОН конец геномной ДНК используется в качестве затравки для обратной транскрипции. Вторая субъединица связывается с геномной ДНК ниже сайта мишени.

(2) Конформационные изменения первой субъединицы позволяет позиционировать домен обратной транскриптазы напротив одноцепочечного разреза геномной ДНК. Обратная транскриптаза, используя ДНК затравку, синтезирует ДНК фрагмент (минус цепь) по РНК матрице.

(3) Вторая субъединица производит разрез второй цепи геномной ДНК.

(4) Последний этап включает в себя синтез второй цепи ДНК по матрице РНК. К настоящему моменту не установлено, какие ферменты осуществляют

синтез второй цепи ДНК: ферменты репарации клетки или один из ферментов элемента R2 (Рисунок 1.7).

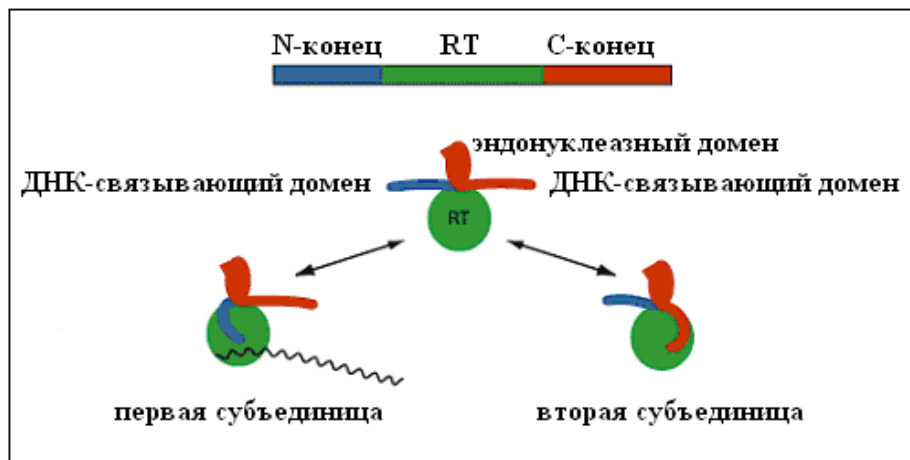
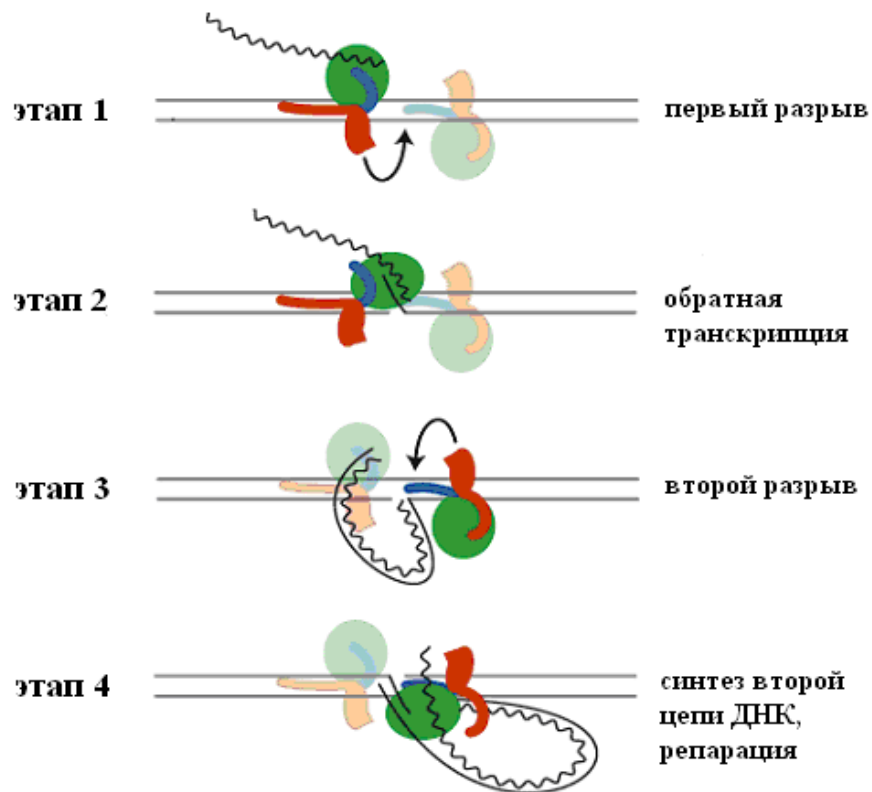


Рисунок 1.7. Модель TPRT реакции REL-endo non-LTR ретротранспозонов (согласно Christensen, Eickbush, 2005). В данной реакции принимают участие две субъединицы белкового комплекса, ярким цветом выделена субъединица активная на определенном этапе. Структура открытой рамки считывания и комплекса ферментов R2 элемента представлена внизу рисунка.

Ферментативный комплекс элемента самостоятельно удаляет РНК матрицу, после завершения синтеза ДНК одноцепочечные разрывы сшиваются, тем самым, завершая встройку элемента (Christensen, Eickbush, 2005). Для сшивки разрывов, по-видимому, используются ферменты клетки-хозяина, так как ни один из

ферментов non-LTR ретротранспозонов не обладает лигазной активностью (Morgan, Gilbert, 2002).

Основные этапы модели TPRT реакции описанной для R2 элемента выполняются в случае APE non-LTR ретротранспозонов (Christensen, Eickbush, 2005). APE эндонуклеаза, закодированная в структуре элементов данной группы, обладает активностями близкими к активностям эндонуклеазы REL-endo non-LTR ретротранспозонов (Malik *et al.*, 1999). К настоящему моменту не установлена роль продукта ORF1 в TPRT реакции. Исследования элемента L1 показали, что белок ORF1 является обязательным для эффективной транспозиции данного элемента и, обладая высокой РНК-связывающей активностью, выполняет функцию шаперона (Martin, Bushman, 2001).

1.4. Класс II – ДНК транспозоны

ДНК транспозоны (класс II) – МГЭ, не использующие РНК посредник в процессе перемещения (транспозиции). В зависимости от количества цепей ДНК, разрезаемых в процессе транспозиции ДНК, транспозоны подразделяются на два подкласса (Wicker *et al.*, 2007) (Рисунок 1.2). Транспозиция элементов подкласса 2 происходит по механизму схожему с механизмом «копирования - встройки» («copy and paste»), включающему репликацию без внесения двуцепочечных разрывов в донорном и акцепторном сайтах.

В подклассе 1 выделяют два порядка, TIR и CRYPTON, характеризующиеся принципиальными различиями в структуре МГЭ (Рисунок 1.2). Элементы порядка TIR содержат концевые инвертированные повторы и кодируют белок транспозазу, в то время как элементы порядка CRYPTON не содержат TIRs и кодируют белок тирозиновую рекомбиназу. Элементы групп *Helitron* и *Maverick* относятся к подклассу 2 (Wicker *et al.*, 2007) (Рисунок 1.2).

Среди представителей класса ДНК транспозонов элементы порядка TIR наиболее активно изучаются исследователями МГЭ. Интерес к данной группе элементов связан с высокой способностью к горизонтальному переносу (ГП). Около 40% описанных случаев ГП было выявлено в группе TIR ДНК транспозонов (Wallau *et al.*, 2012). Дополнительным стимулом к активному изучению является использование элементов данного порядка в области молекулярной биологии,

генной инженерии и медицине (Ohlfest *et al.*, 2004; Bowers *et al.*, 2006; Ott *et al.*, 2006; Aronovich *et al.*, 2007; Izsvak *et al.*, 2009; Peng *et al.*, 2009; Kalendar *et al.*, 2011; Di Matteo *et al.*, 2012).

1.4.1. Порядок TIR. Общая структура TIR ДНК транспозонов

TIR ДНК транспозоны характеризуются относительно простой структурной организацией. Основными структурными компонентами данных элементов являются ген белка транспозазы и концевые инвертированные повторы, фланкирующие тело элементов (Рисунок 1.8).

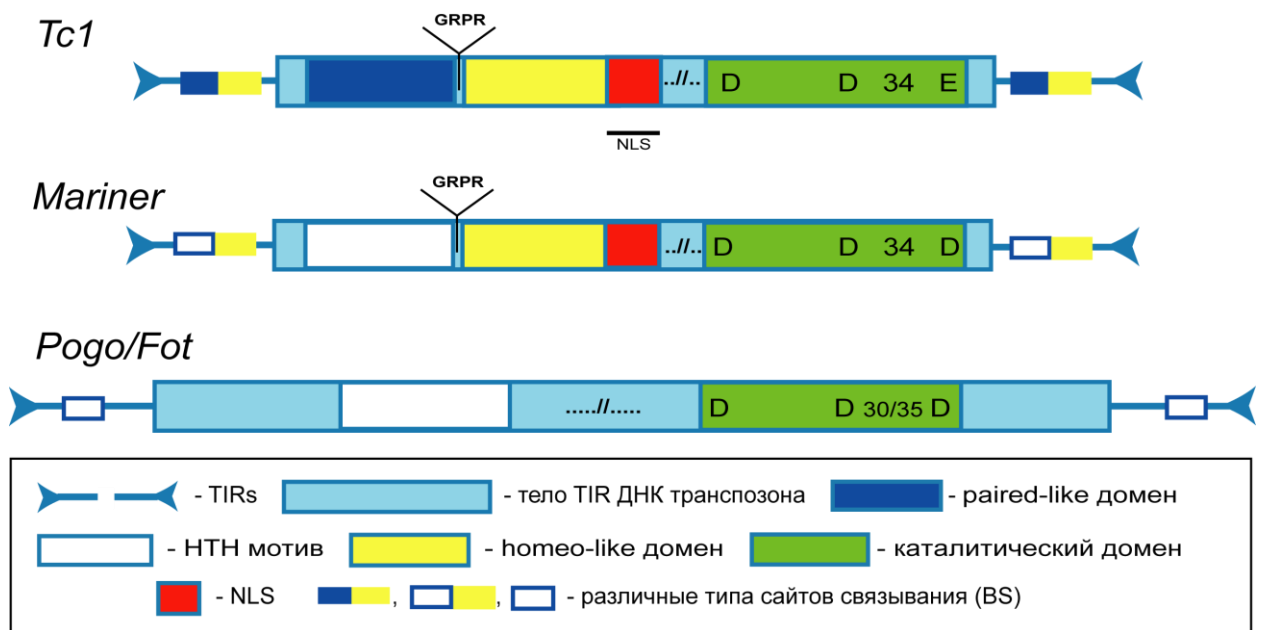


Рисунок 1.8. Структура TIR ДНК транспозонов на примере элементов *Tc1*, *mariner* и *pogo/Fot* (согласно Plasterk *et al.*, 1999). Аббревиатура: BS – binding site (сайт связывания), GRPR – аминокислотная последовательность глицин-аргинин-пролин-аргинин, HTN мотив – helix - turn – helix (мотив спираль-поворот-спираль), NLS – nuclear location signal (сигнал ядерной локализации), TIR – terminal inverted repeat (концевой инвертированный повтор).

Длина и последовательность концевых инвертированных повторов являются специфичными для определенных суперсемейств TIR элементов. Тело элементов, как правило, содержит единственную рамку считывания, кодирующую основной фермент перемещения TIR элементов – транспозазу (Рисунок 1.8).

Транспозаза. Фермент транспозаза распознает последовательности концевых инвертированных повторов, осуществляет вырезание элемента и последующую встройку в новый сайт-мишень. Транспозаза ДНК транспозонов относится к суперсемейству белков ферментов – ретровирусных интеграз.

Структура транспозазы детально изучена на примере представителей наиболее широко распространенного суперсемейства *Tc1/mariner*. В структуре большинства транспозаз порядка TIR выделяют три ключевых домена: ДНК-связывающий, каталитический и NLS (nuclear location signal) (Рисунок 1.8).

N-концевой ДНК-связывающий участок состоит из одного или двух доменов, каждый из которых представлен мотивом типа спираль-поворот-спираль (НТН-helix-turn-helix) (Petrokovski, Henikoff, 1997). Первый домен (paired-like domain) образован первым НТН мотивом и обладает сходством со связывающими доменами некоторых транскрипционных факторов (Ivics *et al.*, 1996). Аминокислотная последовательность GRPR (глицин-аргинин-пролин-аргинин), связывающая paired-like домен со вторым НТН мотивом (или homeo-подобным доменом – homeo-like domain), консервативна для транспозаз *Tc1* и *mariner* элементов (Рисунок 1.8).

Домен, содержащий сигнальную последовательность, необходимую для переноса белка в ядро (NLS), частично перекрывается с ДНК-связывающим доменом транспозазы (Ivics *et al.*, 1996) (Рисунок 1.8). Наличие NLS домена позволяет транспозонам использовать рецептор-зависимую транспортную систему клетки-хозяина для проникновения транспозазы в ядро.

Каталитический домен транспозазы отвечает за разрезание цепей ДНК и реакцию присоединения в процессе транспозиции. Данный домен транспозазы характеризуется тройкой аминокислот: двумя аспарагиновыми аминокислотами (D) и третьей глутаминовой (E) или аспарагиновой кислотой (D) (DDE/D мотив или DDE/D box) (Рисунок 1.8). Замена третьей консервативной аминокислоты D из триады на E у представителей отдельных семейств может приводить к полной потере функциональности транспозазы.

Помимо наличия консервативной триады аминокислот, ключевую роль в функционировании фермента играет расстояние между данными аминокислотными остатками, которое может варьировать в различных группах TIR. DDE/D мотив играет ключевую роль в связывании двухвалентных катионов (Mg^{2+} или Mn^{2+}), являющихся кофакторами фермента, и формировании разрывов в донорном сайте ДНК при вырезании элементов (Richardson *et al.*, 2006).

Кроме вышеописанных доменов, ряд TIR элементов могут содержать дополнительные домены, такие как инсерционный домен элементов суперсемейства *hAT* (Zhou *et al.*, 2004; Hickman *et al.*, 2005).

Концевые инвертированные повторы. Наиболее просто организованные группы TIR ДНК транспозонов (*Tc1* и *mariner*) характеризуются короткими инвертированными повторами, содержащими по одному сайту связывания транспозазы на повтор (Рисунок 1.8). Более сложно организованные элементы могут содержать по два сайта связывания на повтор (внутренний и внешний), однако зачастую лишь одна пара является критической для перемещения (Plasterk *et al.*, 1999). В случае элементов, ДНК связывающий домен транспозазы которых содержит homeo-подобный и paired-подобный домены, сайты связывания, локализованные на концевых инвертированных повторах, состоят из двух частей (Рисунок 1.8). С 5'-концевой частью сайта связывается homeo-подобный домен, а с 3'-концевой частью paired-подобный домен транспозазы. При наличии одного ДНК связывающего домена транспозазы (homeo-домен) сайты связывания TIR являются непарными и короткими (12 п.о. у элемента *pogo*) (Wang *et al.*, 1999) (Рисунок 1.8).

1.4.2. Классификация TIR ДНК транспозонов

В основу классификации TIR ДНК транспозонов положены различия в последовательности транспозазы, длине и последовательности TIRs и размере TSDs, формируемых при встраивании копии элементов в новый сайт-мишень. Согласно современной классификации порядок TIRs ДНК транспозонов включает в себя 17 суперсемейств *hAT*, *Tc1/mariner*, *Merlin*, *Transib*, *PiggyBac*, *Zator*, *Academ*, *Kolobok*, *P*, *Novosib*, *Ginger*, *PIF/Harbinger*, *MULE (Mutator-like)*, *CMC*, *Solo1*, *Solo2* и *Solo3* (Yuan, Wessler, 2011) (Рисунок 1.9).

Среди вышеописанных групп самыми распространенными и разнообразными элементами являются представители суперсемейства *Tc1/mariner*. Первые представители данного суперсемейства были обнаружены в геноме *Drosophila mauritiana*, а к настоящему моменту были выявлены в геномах плоских червей, кишечнополостных, насекомых, растений, млекопитающих и человека (Robertson, 1993; Feschotte, Wessler, 2002; Arkhipova, Meselson, 2005).

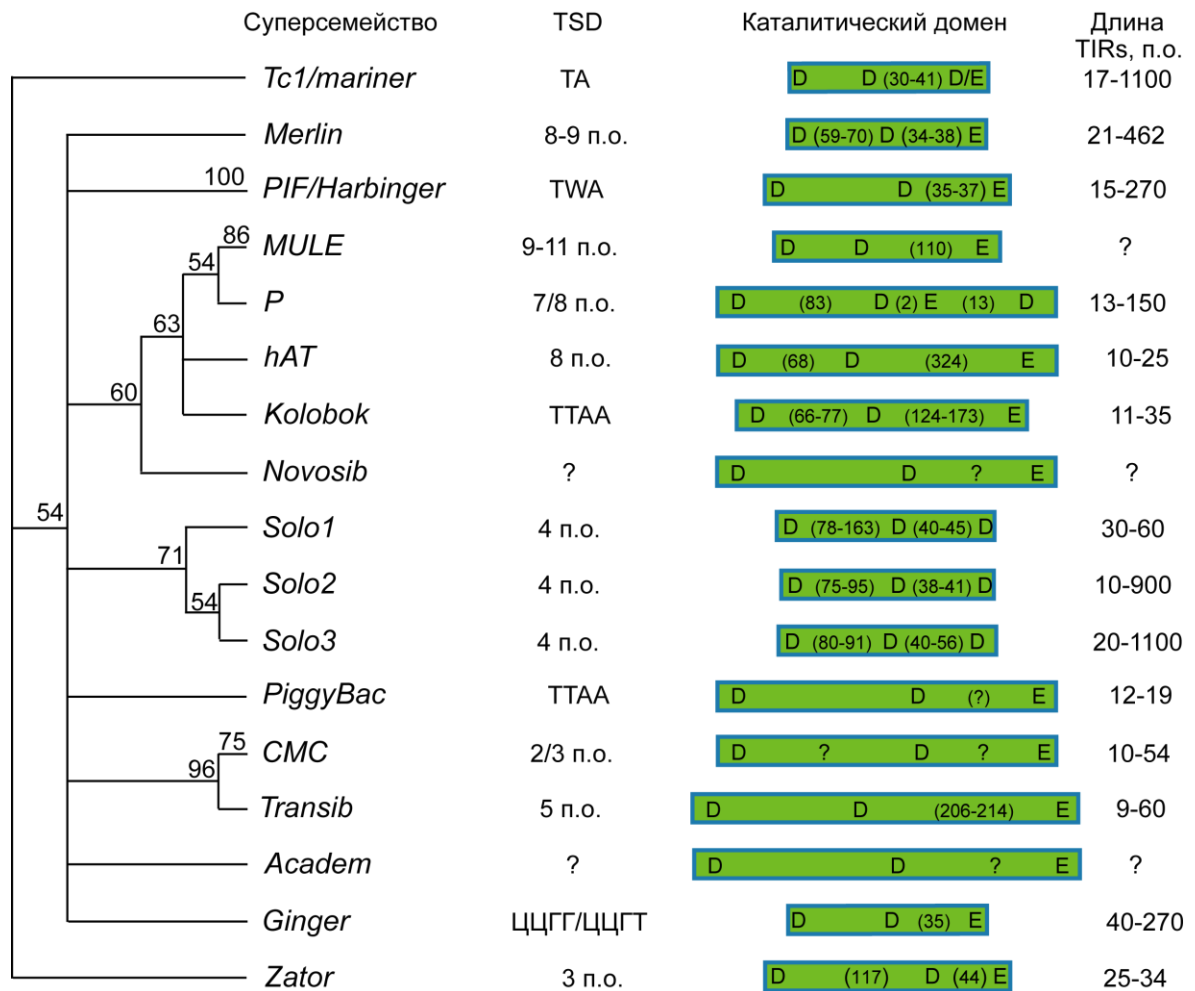


Рисунок 1.9. Дендрограмма суперсемейств ДНК транспозонов порядка TIR, реконструированная на основе последовательностей каталитического домена транспозазы (согласно Yuan, Wessler, 2011). Справа приведена схема строения каталитического домена и отражены основные особенности каждого суперсемейства (TSD и длина TIRs). Для DDE/D триады указаны расстояния между консервативными аминокислотами (в аминокислотных остатках). Аббревиатура: п.о. – пара оснований; TIRs - terminal invert repeats (концевые инвертированные повторы); TSD - target site duplication (дубликация сайта - мишени); W – обозначает наличие нуклеотида А или Т в последовательности.

Транспозоны суперсемейства *Tc1/mariner* являются относительно короткими мобильными элементами (от 1000 до 5000 п.о.) с концевыми инвертированными повторами, варьирующими по длине от 17 п.о. до 1100 п.о. (Plasterk *et al.*, 1999; Leroy *et al.*, 2003). Данные элементы кодируют единственный белок – транспозазу, размер которой варьирует от 282 до 345 а.о.. В данном суперсемействе присутствуют элементы, содержащие как DDD-box, так и DDE-box в каталитическом домене транспозазы. Первый и второй аминокислотные остатки триады разделены 92 а.о., а расстояние между вторым остатком и третьим может варьировать от 30 до 41 а.о. (Brillet *et al.*, 2007).

В суперсемействе *Tc1/mariner* выделяют 7 семейств: *maT*, DD37E, *Tc1-like*, *Gambol*, DD39D, *mariner-like*, *ludens* (Rouault *et al.*, 2009; Munoz-Lopez, Garcia-Perez, 2010). Семейства в составе суперсемейств TIR ДНК транспозонов имеют сложную структуру. Так, современной классификации в составе семейства *mariner-like* (DD34D) выделяют 15 подсемейств (Rouault *et al.*, 2009). Из 15 подсемейств семейства *mariner*, 6 имеют структуру клана (*DTTMarBRI* (*briggsae*), *DTTMarCAP* (*capitata*), *DTTMarCRI*, *DTTMarELE* (*elegans*), *DDTMarIRR* (*irritans*), *DTTMarROS* (*rosa*)), 4 представляют собой серии кланов (*DTTMarAtl* (*atlantis*), *DTTMarCec* (*cecropia*), *DTTMarMel* (*mellifera*), *DTTMarMau* (*mauritiana*)). Оставшиеся 5 групп *DTTMarLin* (*lineata*), *DTTMarUrt*, *DTTMarVer* (*vertumana*), *DTTMar* и *DTTMarGGS* имеют статус подсемейств (Rouault *et al.*, 2009). Элементы различных подсемейств, как правило, выявлены в определенной группе организмов-хозяев. Так, подсемейство *DTTMarMau* образовано элементами только из геномов насекомых отряда *Diptera*, а клан *DTTMarBRI* образован элементами из геномов *C. elegans* и *C. briggsae* (Witherspoon, Robertson, 2003).

1.4.3. Разнообразие и распространение TIR ДНК транспозонов

Некоторые суперсемейства TIR ДНК транспозонов имеют широкое распространение среди различных групп живых организмов (Рисунок 1.10).

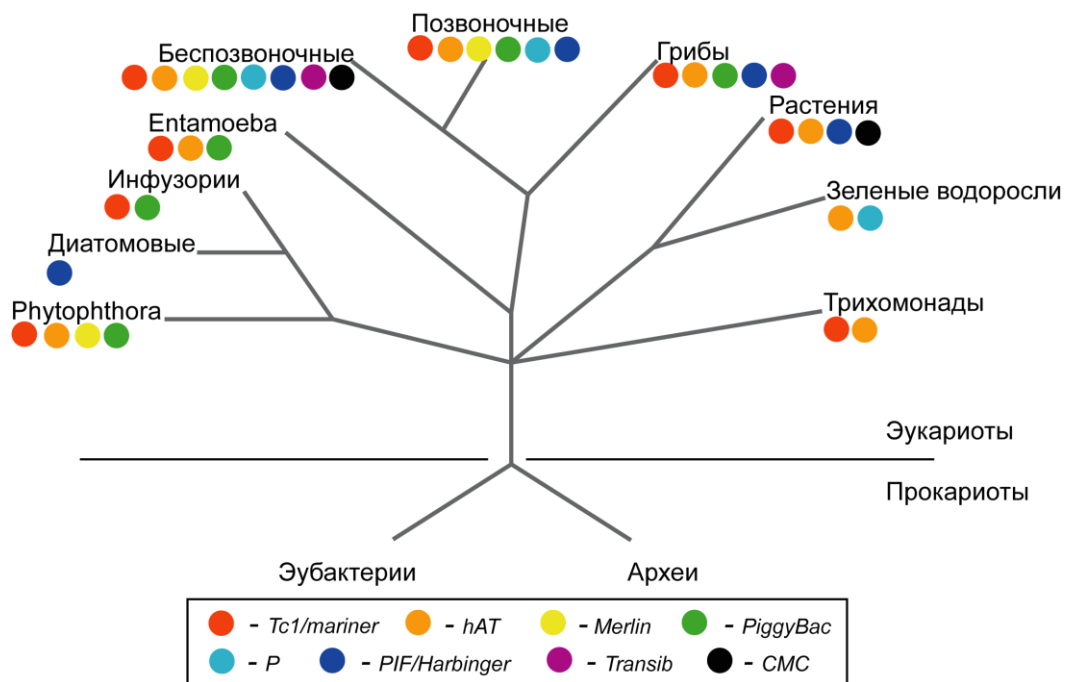


Рисунок 1.10. Распространение основных семейств TIR ДНК транспозонов в различных группах живых организмов (согласно Feschotte, Pritham, 2007).

Так, суперсемейство *hAT* является крупным и древним суперсемейством TIR ДНК транспозонов, появившимся в геномах эукариот задолго до дивергенции растений, животных и грибов (Feschotte, Pritham, 2007) (Рисунок 1.10). Другие суперсемейства, напротив, распространены лишь в ограниченных группах живых существ.

Так, длительное время считалось, что представители группы *CASTA* TIR ДНК транспозонов (суперсемейство *СМС*) распространены исключительно в геномах растительных организмов. Впоследствии, элемент *SmTRC1* данной группы был обнаружен в геноме паразита *Schistosoma mansoni* (DeMarco *et al.*, 2006) (Рисунок 1.10).

1.4.4. Перемещение TIR ДНК транспозонов (транспозиция)

Транспозиция TIR ДНК транспозонов – нерепликативный процесс перемещения копий элементов из донорного в акцепторный сайт мишень по механизму «вырезания-встройки». Данный процесс детально изучен для элементов суперсемейства *Tc1/mariner* и включает четыре основных этапа (Munoz-Lopez, Garcia-Perez, 2010):

(1) На первом этапе две молекулы транспозазы распознают последовательность концевых инвертированных повторов элемента, связываются с ними через ДНК-связывающие домены (НТН мотивы) и формируют Single-End Complex (SEC).

(2) На втором этапе оба фермента формируют одноцепочечный разрыв с 5'-концов TIRs путем гидролиза фосфодиэфирной связи.

(3) На третьем этапе происходит связывание двух молекул фермента, что приводит к сближению TIRs и образованию Paired-End Complex (PEC). На данной стадии фермент участвует в процессе в форме димера. Затем происходит гидролиз фосфодиэфирных связей с 3'-конца TIRs. Двухцепочечный разрыв в донорном сайте репарируется за счет ферментов клетки-хозяина.

(4) На завершающем этапе PEC связывается с ДНК в области сайта мишени и формирует так называемый комплекс захвата мишени (Target Capture Complex - TCC). Сайтом-мишенью в случае *Tc1/mariner* элементов может являться любой ТА

динуклеотид. 5'-конец сайта мишени подвергается нуклеофильной атаке со стороны свободной 3'-ОН группы трансозона, образовавшейся в результате вырезания элемента. Разрывы сшиваются за счет ферментативной системы клетки-хозяина, в результате чего формируются TSDs фланкирующий встроенный трансозон (Рисунок 1.11).

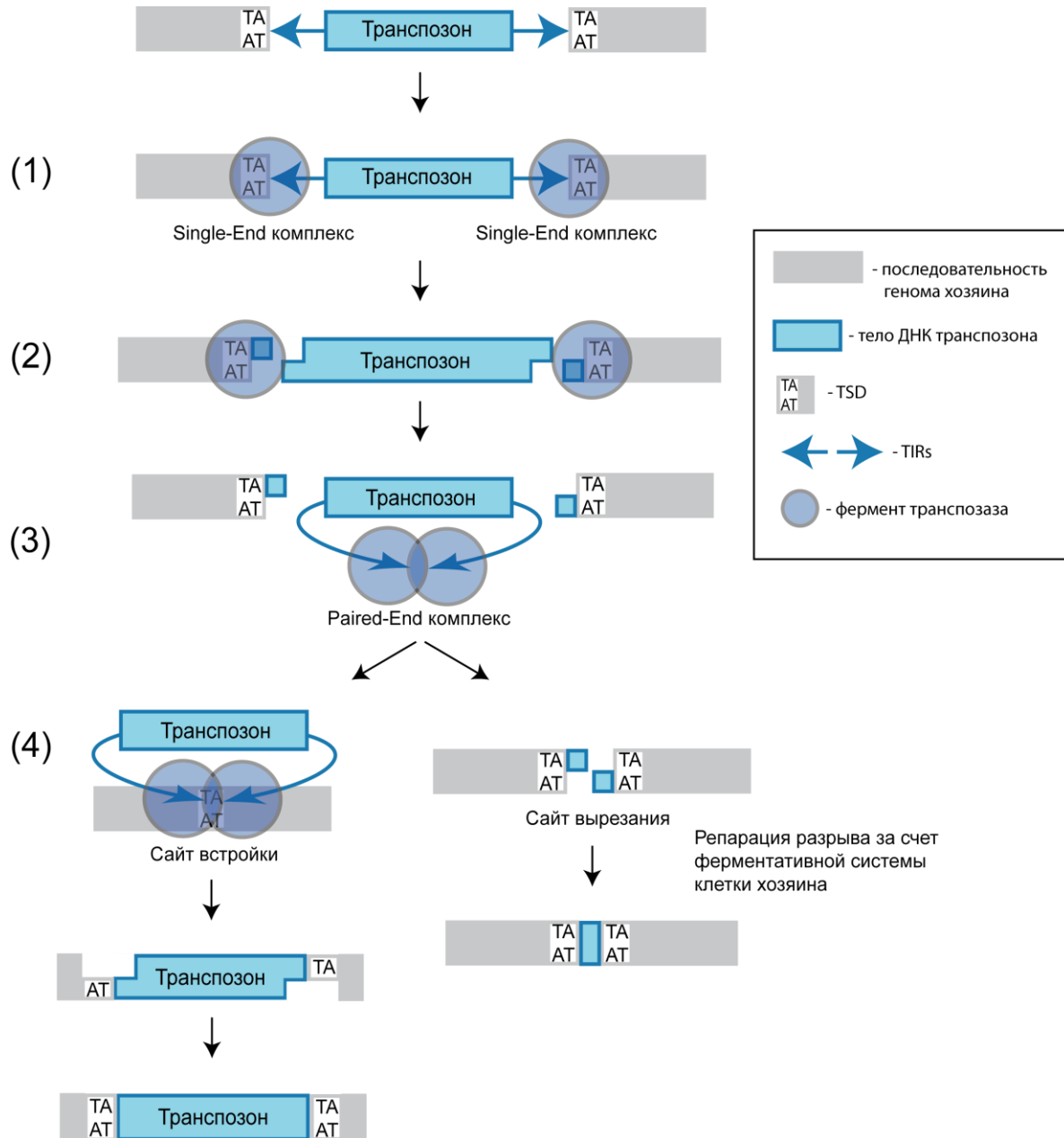


Рисунок 1.11. Основные этапы «cut and paste» транспозиции на примере представителей суперсемейства *Tc1/mariner* ДНК трансозонов (согласно Munoz-Lopez, Garcia-Perez, 2010). Аббревиатура: TIRs - terminal invert repeats (концевые инвертированные повторы); TSD - target site duplication (дупликация сайта – мишени).

Для осуществления вышеописанного процесса не требуется дополнительная энергия АТФ, поскольку энергия необходимая для формирования фосфодиэфирной

связи выделяется в процессе разрезания ДНК в сайте встройки (Munoz-Lopez, Garcia-Perez, 2010).

Для вырезания и встройки транспозона ферменту транспозазе требуется только двухвалентный катион (Mg^{2+} или Mn^{2+}) (Liu *et al.*, 2007). Отсутствие зависимости транспозиции *Tc1/mariner* от специфических ферментов и факторов клетки-хозяина делает данные элементы идеальными инструментами трансгеноза.

1.5. Жизненный цикл мобильных генетических элементов

МГЭ представляют собой эгоистическую ДНК, основной целью которой является копирование и распространение собственных копий в геноме хозяина. Жизненный цикл МГЭ представлен на рисунке 1.12.



Рисунок 1.12. Жизненный цикл мобильных генетических элементов (согласно Schaack *et al.*, 2010; Munoz-Lopez, Garcia-Perez, 2010).

Жизненный цикл состоит из следующих этапов: появления нового активного элемента в организме-хозяине, колонизация генома зародышевой линии, увеличение числа копий элемента и распространение в популяции, диверсификация, инактивация копий за счет накопления мутаций (вертикальная инактивация) или негативного отбора, исчезновение неактивных копий элемента из генома (Schaack *et al.*, 2010) (Рисунок 1.12).

Наиболее эффективной для распространения и сохранения МГЭ в ряде поколений является перемещение в клетках зародышевой линии. Напротив, транспозиция МГЭ в соматических клетках не является продуктивной с точки зрения эволюции, так как новые копии будут утеряны с гибелью организма-хозяина (Han, 2010).

Организм-хозяин выработал механизмы защиты против распространения МГЭ в геноме, ввиду негативного воздействия встроок МГЭ в определенных сайтах, эктопической рекомбинации между МГЭ и метаболических затрат на «содержание» МГЭ. Существует два основных способа избежать исчезновения МГЭ (выйти из вышеописанного цикла): снижения негативного воздействия МГЭ на геном и перенесение элемента в геном нового хозяина (горизонтальный перенос) (Schaack *et al.*, 2010).

1.5.1. Снижение негативного воздействия МГЭ на геном хозяина

Успешное распространение транспозона зависит от активной транспозиции в клетках зародышевой линии. Однако, чрезмерно высокий уровень транспозиции, оказывает потенциально негативное воздействие на геном хозяина и на сам транспозон, чей «успех» также зависит и от выживаемости хозяина (Munoz-Lopez, Garcia-Perez, 2010). Таким образом, выработка стратегий снижения негативного воздействия МГЭ на геном путем стабилизации числа копий элемента является выгодной как для организма-хозяина, так и для МГЭ. Стратегии снижения негативного воздействия МГЭ на геном можно разделить на два типа: саморегуляция и регуляция, осуществляемая организмом-хозяином (Blumenstiel, 2011).

Саморегуляция. Саморегуляция предполагает снижение уровня транспозиции до определенного значения. Примером данного типа регуляции является ингибирование гиперпродукции (overproduction inhibition - OPI) описанное для *Tc1/mariner* ДНК транспозонов (Lohe, Hartl, 1996). Фермент транспозаза данной группы элементов выступает в качестве ингибитора процесса транспозиции. При достижении концентрации транспозазы порогового значения – активность транспозиции снижается. Предполагаемым механизмом снижения активности является формирование неактивных димеров транспозазы. При

увеличении количества копий *Tc1/mariner* ДНК транспозонов в геноме возрастает продукция транспозазы, что путем ОРІ снижает уровень транспозиции элементов (Lohe, Hartl, 1996; Lampe *et al.*, 1998).

Регуляция транспозиции организмом-хозяином. Данный тип регуляции транспозиции может осуществляться на стадии транскрипции, трансляции, синтеза новой копии, интеграции в геном. Основные механизмы регуляции транспозиции: транскрипционный сайленсинг, пост-транскрипционный сайленсинг и модификация хроматина (Slotkin, Martienssen, 2007).

1.5.2. Горизонтальный перенос

1.5.2.1. Понятие «горизонтальный перенос»

Горизонтальный перенос (ГП) – процесс передачи генетической информации между репродуктивно изолированными видами. Объектом горизонтального переноса может являться любая генетическая информация – гены, регуляторные последовательности, МГЭ. Большинство случаев горизонтального переноса выявленных между эукариотическими организмами приходится на ГП МГЭ. Способность МГЭ к перемещению и интеграции в новые сайты генома делает их наиболее удобными объектами ГП по сравнению с другими кодирующими и некодирующими последовательностями генома (Schaack *et al.*, 2010).

К настоящему моменту сформулированы три критерия горизонтального переноса (Silva *et al.*, 2004):

- 1) неравномерное распределение объекта ГП между видами определенной таксономической группы;
- 2) высокое сходство последовательностей из геномов репродуктивно изолированных видов, превышающее ожидаемый уровень сходства с учетом времени дивергенции видов хозяев.
- 3) несоответствие филогении построенной на основе объекта ГП и филогении видов хозяев.

1.5.2.2. Горизонтальный перенос МГЭ между эукариотическими организмами

Первым из задокументированных случаев ГП является перенос *P* элемента между представителями рода *Drosophila* - *D. melanogaster* и *D. paulistorum*. (Daniels *et al.*, 1984). К настоящему моменту выявлено 330 случаев горизонтального переноса МГЭ между эукариотическими организмами. Основная масса случаев ГП была выявлена в царстве животных (94.37%) и менее значительное количество случаев в царствах растений (4.30%) и грибов (4.30%) (Wallau *et al.*, 2012) (Рисунок 1.13).

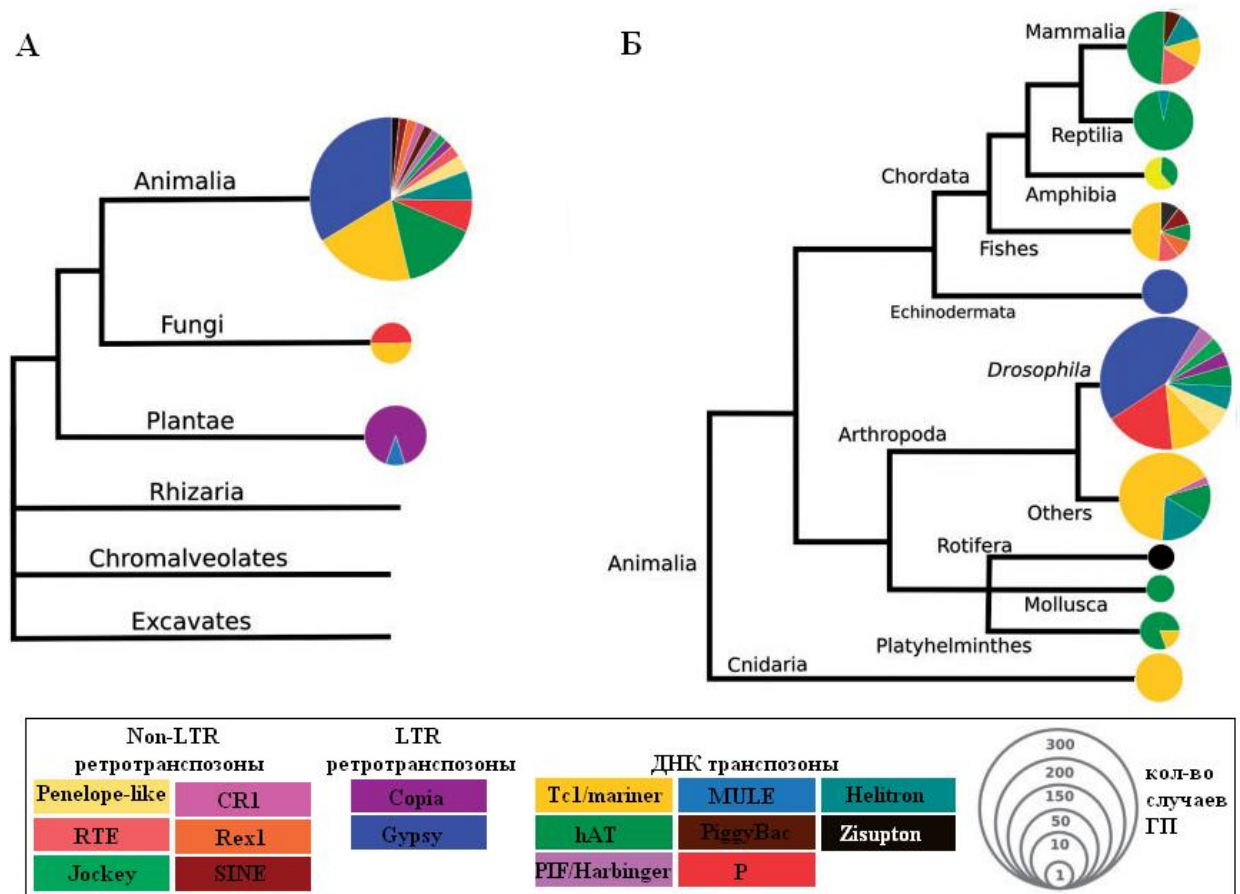


Рисунок 1.13. Распределение случаев горизонтального переноса различных групп МГЭ в геномах эукариотических организмов (А) и геномах представителей царства Животных (Б) (согласно Wallau *et al.*, 2012). Для групп МГЭ не представленных на данном рисунке на настоящий момент не выявлено случаев горизонтального переноса.

Различие в частоте встречаемости ГП между царствами может объясняться различной восприимчивостью организмов к переносу, меньшим количеством геномных проектов представителей царства Растений по сравнению с царства Животных и Грибов, наиболее активным использованием животных моделей для

исследования биологических процессов. Так 178 из 330 случаев ГП было выявлено в геномах представителей рода *Drosophila* (54%) (Loreto *et al.*, 2008).

Основное количество случаев ГП описано для ДНК транспозонов и LTR ретротранспозонов, а на non-LTR ретротранспозоны приходится не более 5% случаев (Silva *et al.*, 2004; Loreto *et al.*, 2008; Schaack *et al.*, 2010). Неравномерное распределение случаев горизонтального переноса между группами МГЭ связано с особенностями структуры и функционирования элементов данных групп. Эффективность горизонтального переноса, по-видимому, определяется:

- 1) стабильной интеграцией МГЭ в геном, для которой требуется образование стабильного внехромосомного интермедиата в процессе перемещения;
- 2) эффективным увеличением числа копий в новом геноме, которое определяется наличием или отсутствием потребности в ферментах и факторах организма-хозяина для транспозиции МГЭ и наличием механизмов саморегуляции транспозиции (Silva *et al.*, 2004).

ДНК транспозоны обладают рядом преимуществ для успешного горизонтального переноса: относительно простое строение; наличие стабильных интермедиатов в процессе перемещения; автономность процесса перемещения (транспозиция данных элементов не зависит от ферментативной системы хозяина); обладают механизмами саморегуляции активности транспозиции, что позволяет эффективно сохраняться в новом геноме (Lohe, Hartl, 1996; Wicker *et al.*, 2007; Munoz-Lopez, Garcia-Perez, 2010; Schaack *et al.*, 2010).

Вышеописанные свойства позволяют ДНК транспозонам переноситься горизонтально между эволюционно удаленными видами эукариот с высокой эффективностью. Из 153 описанных случаев горизонтального переноса 25.49% произошли между представителями разных царств, 3.26% между различными классами, 17.54% между порядками, 11.76% между семействами, 3.26% между родами, и 37.90% случаев являются ГП между представителями одного рода (Wallau *et al.*, 2012).

LTR ретротранспозоны характеризуются наличием стабильного интермедиата в процессе перемещения и способностью к самостоятельному

горизонтальному переносу (Wallau *et al.*, 2012). Кроме того, LTR ретротранспозоны эволюционно близки к группе ретровирусов и могут обладать способностью синтезировать белок оболочки вирусных частиц (ENV) (Kim *et al.*, 1994; Song *et al.*, 1994). Инфекционная способность LTR ретротранспозонов является основной характеристикой определяющей эффективный горизонтальный перенос данной группы элементов.

В процессе транспозиции non-LTR ретротранспозоны образуются РНК интермедиат, который подвергается обратной транскрипции в ядре, непосредственно вблизи сайта будущей встройки (Christensen, Eickbush, 2005). В данном случае не формируется стабильный внехромосомный ДНК интермедиат, как в случае ДНК транспозонов и LTR ретротранспозонов. Согласно мнению большинства исследователей, отсутствия возможности стабильного переноса в новый геном достаточно для объяснения низкого уровня ГП non-LTR ретротранспозонов.

1.5.2.3. Механизм горизонтального переноса

Механизм ГП МГЭ между различными организмами изучен слабо. Для успешного ГП требуется доставка МГЭ от донора к новому хозяину или в линию клеток зародышевого пути (в случае многоклеточного организма) и интеграция в геном. Для передачи МГЭ требуется наличие вектора, защищающего ДНК и РНК от деградации и осуществляющего передачу материала, и наличие контакта между геномом донором и геномом реципиентом. Некоторые LTR ретротранспозоны способны самостоятельно формировать вирусоподобные частицы, которые могут служить в качестве векторов ГП. В случае других МГЭ в качестве векторов могут выступать внутриклеточные паразиты эукариот, такие как бактерии и вирусы. Данные векторы могут проникать и покидать клетку и взаимодействовать с геномом. Основное препятствие для использования подобных векторов заключается в низкой вероятности фиксации и сохранения чужеродной ДНК в геномах вирусов и бактерий.

Вирусные векторы. Существует ряд примеров переноса МГЭ из генома эукариотического организма в геном вируса. Для ДНК транспозонов *piggyBac* и LTR ретротранспозона *TE1* в лабораторных условиях был показан перенос из

геномов насекомых отряда Lepidoptera в генома бакуловирусов (Friesen, Nissen, 1990; Jehle *et al.*, 1998). В естественных условиях был показан ГП non-LTR ретротранспозона из генома змеи в геном поксвируса (Piskurek, Okada, 2007).

Внутриклеточные паразиты. Другим потенциальным вектором могут являться бактерии рода *Wolbachia*, поражающие насекомых и нематод. Бактерии данного рода активно взаимодействуют с геномом хозяина, ряд генов *Wolbachia* был обнаружен в геноме насекомых-хозяев (Dunning Hotopp *et al.*, 2007). Аналогичная возможность переноса собственного генетического материала в геном хозяина была показана для трипаносом (Hecht *et al.*, 2010).

Эктопаразиты. Предполагаемым вектором ГП МГЭ между насекомыми могут являться эктопаразитические клещи, например представители вида *Proctolaelaps regalis*. Не смотря на то, что последовательности МГЭ не были выявлены в геноме данного паразита, для функционирования в качестве вектора не требуется его интеграция в геном (Houck *et al.*, 1991). Геном паразита триатомового клопа вида *Rhodnius prolixus* содержит четыре семейства ДНК транспозонов, элементы которых представлены в геномах позвоночных животных, на которых данный клоп паразитирует (Gilbert *et al.*, 2010).

1.5.2.4. Влияние горизонтального переноса на МГЭ и геном хозяина

ГП является ключевым этапом жизненного цикла МГЭ, позволяющим элементам избежать инактивации, стохастической потери и колонизировать новые геномы (Venner *et al.*, 2009) (Рисунок 1.12). ГП МГЭ является источником нового генетического материала и модификатором архитектуры генома хозяина. Массовый ГП генетического материала может привести к значительному увеличению размеров генома. Так, в геноме летучей мыши *Myotis lucifugus* представлены ДНК транспозоны четырех различных семейств (общей длиной в 21 млн.п.о.), которые были перенесены в геном путем ГП около 30 млн. лет назад (Rase *et al.*, 2008). Кроме того, ГП может привести к появлению новых генов, изменению существующих генов и регуляторных последовательностей. Например, ген транспозазы из семейства ДНК транспозонов SPACE INVADERS, был захвачен

геномом реципиентом и формирует новый смешанный ген специфичный для мышевидных грызунов (Pace *et al.*, 2008).

1.6. Влияние мобильных генетических элементов на геном хозяина

МГЭ составляют значительную фракцию геномов эукариотических организмов и оказывают существенное влияние на эволюцию геномов хозяев. Благодаря уникальным свойствам различные типы МГЭ оказывают различное влияние на геном хозяина (Feschotte, Pritham, 2007). Так, эффект от встройки ДНК транспозонов может являться временным, поскольку при транспозиции из донорного сайта вырезается исходная копия. Встройки же ретротранспозонов, как правило, являются постоянными, поскольку при ретротранспозиции не происходит вырезание исходного элемента. МГЭ оказывают влияние на геном хозяина путем изменения состава генов, функций генов, структуры и размера генома.

1.6.1. Воздействие МГЭ на гены генома хозяина

Характер воздействия встройки МГЭ на ген-мишень зависит от позиции встройки. Первым вариантом является встраивание МГЭ непосредственно в кодирующую область гена. Наиболее очевидным результатом такой встройки МГЭ является прямое нарушение кодирующей части гена, что приводит к ингибированию образования продукта данного гена. Однако, встройка МГЭ в геном может повлиять на экспрессию генов различными способами. МГЭ могут содержать в своей структуре сигналы полиаденилирования, в данном случае встройка МГЭ приведет к формированию внутригенного сигнала полиаденилирования и преждевременной терминации транскрипции (Roy-Engel *et al.*, 2005). Другим примером воздействия встройки МГЭ является «экзонизация» - явление образования нового экзона из последовательности МГЭ. Так, С – концевой экзон рецептора лептина 219.1 у человека является примером нового экзона сформированного из последовательности неавтономного элемента *SINE-R* (Ostertag *et al.*, 2003).

Вторым вариантом является встраивание МГЭ в промоторную или энхансерную область гена. Данное явление широко распространено в различных группах эукариот. Практически 25% промоторных районов генов человека

содержат последовательности, произошедшие из МГЭ (Jordan *et al.*, 2003). Встройка МГЭ в энхансерную области генов может приводить к развитию различных заболеваний человека (Chen *et al.*, 2005).

1.6.2. МГЭ и изменения размеров генома

Распространение МГЭ является причиной изменения размеров геномов растений, животных и простейших, поскольку МГЭ является источником кодирующего и не кодирующего генетического материала. Основной вклад в изменение размеров геномов вносят ретротранспозоны. Однако ДНК транспозоны также могут вносить вклад в изменение размеров геномов, например у *Trichomonas vaginalis* и *Myotis lucifugus* (Feschotte, Pritham, 2007).

МГЭ оказали значительное влияние на изменение размеров геномов растений. Так, распространение МГЭ привело к формированию значительных различий в размерах геномов представителей различных видов риса (род *Oryza*) (Zuccolo *et al.*, 2007). Количество МГЭ в геномах животных также существенно варьирует, от 77% в случае генома лягушки *Pelophylax esculentus* до 12% в геноме *C. elegans* (Biemont, Vieira, 2006). Распространение и элиминация МГЭ оказали значительное влияние на формирование новых видов животных, например в случае грызунов. Распространение МГЭ оказало влияние и на изменение размеров геномов простейших. Так, геном оомицета *Phytophthora infestans* является самым крупным в группе хромист и на 74% состоит из МГЭ (Haas *et al.*, 2009). Сравнение трех видов рода *Phytophthora* (*P. infestans*, *P. sojae* и *P. ramorum*) показало, что в геноме *P. infestans*, содержится в 5 раз больше LTR ретротранспозонов по сравнению с геномами *P. sojae* и *P. ramorum* (Haas *et al.*, 2009).

1.6.3. МГЭ и хромосомные перестройки

Помимо изменения размеров геномов, МГЭ влияют на структуру геномов, способствуя хромосомным перестройкам. МГЭ могут способствовать формированию делеций, транслокаций и инверсий в геномах растений, позвоночных и беспозвоночных животных (Chenais *et al.*, 2012). Активность элементов класса I таких как *BEL*, *roo*, *Doc* и *I* и элементов класса II, таких как *Hobo*, *FB* и *P*, связывают с формированием хромосомных перестроек у видов рода

Drosophila (Lim, Simmons, 1994). Элементы *L1* и *Alu*, широко представленные в геноме человека, участвуют в неаллельной гомологичной рекомбинации (non-allelic homologous recombination - NAHR) и, как следствие, являются причиной формирования делеций, дупликаций, инверсий или транслокаций. Кроме того, хромосомные перестройки, вызванные *L1* и *Alu* элементами, являются причиной развития различных заболеваний человека, таких как онкологические. Так транслокация между 9 и 22 хромосомой человека, вызванная *Alu*, получила название Philadelphia, и приводит к развитию хронического миелоидного лейкоза (Jeffs *et al.*, 1998).

1.6.4. Молекулярная domestикация генов МГЭ

Одним из основных вкладов МГЭ в эволюцию генома хозяина является предоставление генетического материала для сборки новых генов и формирования новых функций. Данное явление получило название «молекулярная domestикация». В настоящее время в геномах растений, животных и грибов выявлен ряд генов domestичированных от МГЭ. От 50 до 100 генов млекопитающих и растений произошли от кодирующих последовательностей ДНК транспозонов и ретротранспозонов (Alzohairy *et al.*, 2013).

Domestикация генов ДНК транспозонов. Большинство генов произошедших от ДНК транспозонов кодируют белки близкие к транспозазе. Исключение составляют с-интегразы и белки, содержащие MADF домен, domestичированные от групп *Maverick* и *PIF/Harbinger*, соответственно (Alzohairy *et al.*, 2013).

Ярким примером domestикация генов ДНК транспозонов является происхождение иммунной системы позвоночных (в качестве обзора Feschotte, Pritham, 2007). Между V(D)J рекомбинацией (процесс соматической рекомбинации ДНК генов иммуноглобулинов) и транспозицией ДНК транспозонов наблюдается значительное сходство. Сходство данных процессов связано с высоким сходством ключевого фермента V(D)J рекомбинации RAG1 и транспозазы, кодируемой элементами суперсемейства *Transib*. Сайты связывания данных ферментов имеют высокое сходство на уровне нуклеотидных последовательностей и по пространственной структуре. Кроме того, транспозаза *Transib* и белок RAG1

формируют характерные TSDs длиной 5 п.о. Вышеперечисленные факты подтверждают, что V(D)J рекомбинация является результатом случайной доместикиции транспозазы элемента *Transib*.

Доместикация генов ретротранспозонов. Гены, произошедшие в результате доместикиции ретротранспозонов, кодируют белки близкие по свойствам к интегразе и гену *gag*. Интересным примером доместикиции генов ретротранспозонов является формирование системы поддержания длины теломер (в качестве обзора Alzohairy *et al.*, 2013). Теломеры представляют собой массивы tandemных коротких повторов, присутствующие на концах линейных хромосом эукариот. В результате каждого цикла репликации происходит укорочение длины теломер (феномен концевой недорепликации), а специализированная РНК полимеразы (теломеразы) восстанавливает нормальную длину теломер. Теломеразы это рибонуклеопротеидный фермент, который содержит РНК, выступающую в качестве шаблона для репликации теломер. Высокое сходство последовательностей и активных центров теломеразы эукариот и ферментов кодируемых non-LTR ретротранспозонами было выявлено в работах Nakamura *et al.* (1997) и Malik, Eickbush (2001). Таким образом, обратная транскриптаза non-LTR ретротранспозонов была доместицирована эукариотическими геномами и приобрела функцию сохранения длины теломер.

1.7. Методы изучения мобильных генетических элементов

Биоинформатические методы анализа МГЭ можно условно разделить на несколько групп:

- (1) методы поиска нуклеотидных и аминокислотных последовательностей в базах данных и прочитанных геномах;
- (2) методы сравнительного анализа последовательностей;
- (3) методы филогенетического анализа последовательностей.

Отдельную группу формируют методы анализа горизонтального переноса МГЭ (Jordan, Bowen, 2004) (Рисунок 1.14).

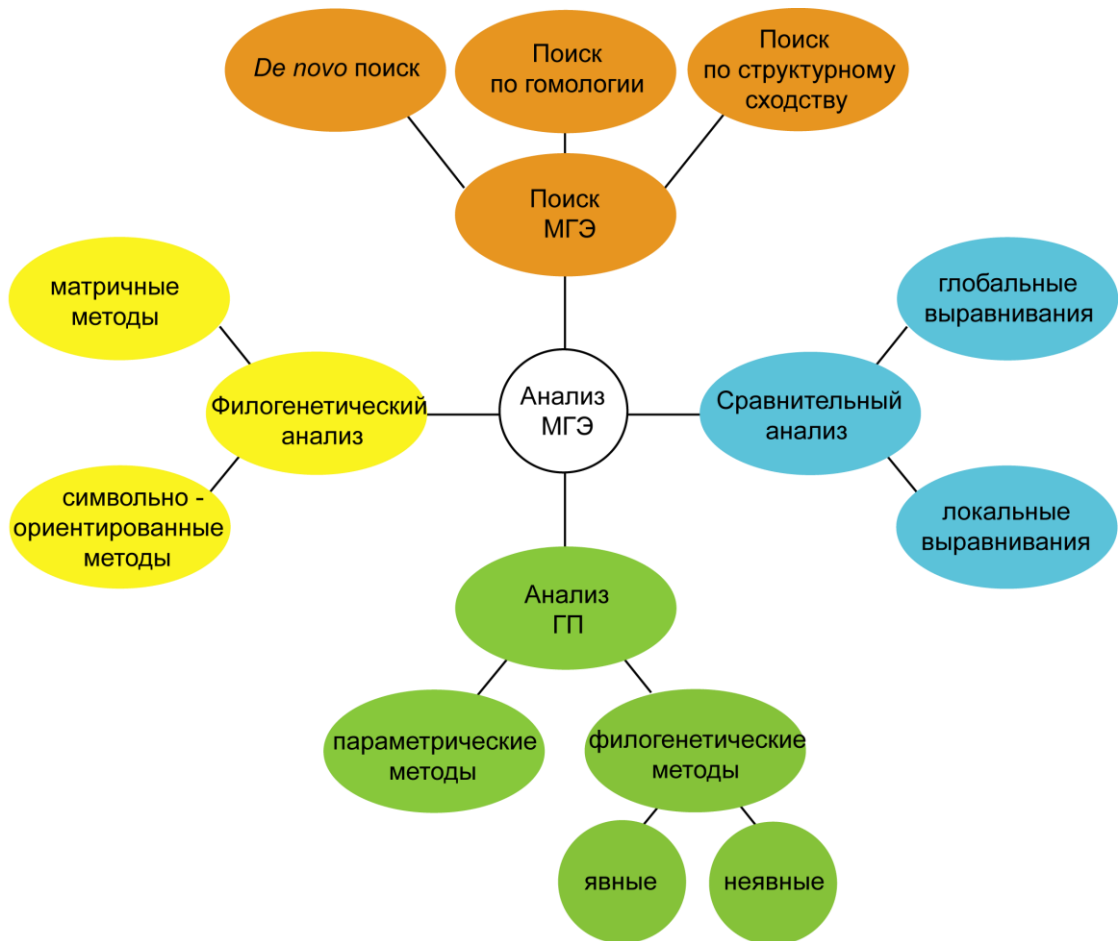


Рисунок 1.14. Схема основных групп методов изучения МГЭ согласно Jordan, Bowen, 2004.

1.7.1. Методы поиска нуклеотидных и аминокислотных последовательностей МГЭ в базах данных и прочитанных геномах

Методы идентификации МГЭ *de novo* позволяют осуществлять поиск повторенных последовательностей в прочитанных геномах без использования данных о структуре или сходстве МГЭ (Bergman, Quesneville, 2007). Наиболее распространенная стратегия *de novo* поиска заключается в обнаружении пар сходных последовательностей в различных позициях генома при сравнении данного генома с самим собой. Большинство программ используют одну из классических стратегий выявления повторов, таких как суффиксные деревья, поиск с помощью матрицы сходства, хэширование (hashing), k-мерный подход. После идентификации повторов проводится кластеризация выровненных последовательностей в семейства МГЭ и отфильтровывание последовательностей не относящихся к МГЭ. *De novo* методы не позволяют осуществлять

специфический поиск последовательностей МГЭ и выявляют различные группы повторенных последовательностей: тандемные повторы, сателлитные повторы, МГЭ. Примерами программ *de novo* поиска являются Reputer, RepeatFinder, ReAS, SSAHA, BLAT, PASH, WindowMasker, RepeatGluer, RepeatMasker, P-Clouds и другие (Kurtz *et al.*, 2001; Ning *et al.*, 2001; Volfovsky *et al.*, 2001; Kent, 2002; Kalafus *et al.*, 2004; Pevzner *et al.*, 2004; Li *et al.*, 2005; Morgulis *et al.*, 2006; Gu *et al.*, 2008).

Методы идентификации МГЭ по гомологии осуществляют поиск МГЭ за счет выявления последовательностей имеющих сходство с белок кодирующими последовательностями известных МГЭ (Bergman, Quesneville, 2007). В отличие от методик *de novo*, методы идентификации по гомологии позволяют выявлять низкокопийные элементы и классифицировать полученные последовательности. Однако они наиболее эффективно осуществляют поиск ранее описанных МГЭ и элементов, которые сохранили достаточно высокую гомологию на уровне белковой последовательности. Данные методики не позволяют выявлять определенные группы МГЭ, в последовательностях элементов которых отсутствует белок кодирующая часть (MITEs и SINEs). Примерами программ поиска по гомологии являются BLAST, FASTA и HMMER (Pearson, Lipman, 1988; Altschul *et al.*, 1990; Eddy, 2011).

Методы поиска МГЭ по структурному сходству основаны на выявлении специфических участков структуры элементов, таких как длинные концевые повторы в случае LTR ретротранспозонов (Bergman, Quesneville, 2007). Как и методы поиска по гомологии группа методов поиска по структурному сходству позволяет выявлять низкокопийные элементы и классифицировать последовательности МГЭ, однако, недостатком данной группы является необходимость разработки отдельного метода под отдельную группу МГЭ. Кроме того, некоторые группы МГЭ менее не имеют организованной структуры, что не позволяет выявлять их с помощью данных методов. Примерами данной группы методов являются LTR_STRUC, LTR_FINDER, SINEDR, RTAnalyzer, FINDMITE, MUST, MITE Analysis Kit (МАК), HelitronFinder и другие (Tu, 2001; Yang, Hall, 2003; McCarthy, McDonald, 2003; Tu *et al.*, 2004; Lucier *et al.*, 2007; Xu, Wang, 2007; Du *et al.*, 2008; Chen *et al.*, 2009).

1.7.2. Методы сравнительного анализа последовательностей МГЭ

Для сравнительного анализа МГЭ проводят построение выравниваний гомологичных остатков в группе последовательностей. Определение гомологичных участков группы последовательностей является необходимым условием извлечения биологически значимой информации, позволяющей проводить реконструкцию эволюционной истории последовательностей (филогения), предсказание функции каждого остатка, моделирование 3D - структуры, дизайна праймеров для ПЦР анализа и др.

В зависимости от количества анализируемых последовательностей выравнивания подразделяются на парные и множественные. Компьютерные подходы выравнивания последовательностей подразделяются на две группы: глобальные и локальные (Рисунок 1.14). Глобальные подходы выравнивают каждый остаток в любой последовательности анализируемого набора и применяются, в случае если последовательности в наборе обладают высоким сходством и имеют одинаковую длину (пример алгоритм Нидлмана-Вунша). Локальные выравнивания, как правило, применяются для более разнородных групп последовательностей. Данный тип выравниваний на первом этапе определяет и выравнивает районы с высоким сходством и затем выравнивает оставшуюся часть последовательностей (пример алгоритм Смита-Ватермана).

В настоящее время разработана целая группа алгоритмов выравнивания последовательностей, включающая медленные алгоритмы динамического программирования, а также быстрые эвристические алгоритмы или вероятностные методы. Примерами современных методов построения множественных выравниваний являются: ClustalW - наиболее широко используемый (Thompson *et al.*, 1994); Muscle - быстрый и точный (Edgar, 2004); T-COFFEE - значительно более точный, чем два вышеописанных, но существенно более медленный (Notredame *et al.*, 2000).

1.7.3. Методы филогенетического анализа последовательностей МГЭ

Методы филогенетического анализа можно разделить на две группы: методы, основанные на оценке расстояний или матричные методы (distance-based methods), и символно-ориентированные методы (character-based methods) (Jordan,

Bowen, 2004) (Рисунок 1.14). Первая группы методов вычисляет попарные расстояния между последовательностями и строит дерево, в котором расстояния между вершинами наилучшим образом соответствуют матрице попарных расстояний. Наиболее широко используемыми матричными методами являются *алгоритм попарной группировки* (Unweighted Pair Group Method using arithmetic Averages - UPGMA) и *метод соединения ближайших соседей* (Neighbor-joining - NJ). Алгоритм NJ является модификацией алгоритма UPGMA и основан на принципе поиска пар ближайших последовательностей таким образом, чтобы суммарная длина ветвей была минимальной на каждом этапе кластеризации.

Символьные методы используют попарные расстояния с учетом распределения символов в последовательностях для получения оптимальной топологии дерева (Jordan, Bowen, 2004). Наиболее активно используемыми символьными методами являются метод максимальной экономии (Maximum Parsimony, MP) и метод максимального правдоподобия (Maximum likelihood, ML). *Метод максимальной экономии* наиболее популярный алгоритм восстановления наследственных связей. Алгоритм конструирует множество возможных филогенетических деревьев, и затем осуществляет их оценку на основе числа эволюционных изменений (мутаций). Наиболее релевантным является дерево, которое предполагает наименьшее число мутаций для описания эволюционных взаимоотношений между последовательностями. *Метод максимального правдоподобия* конструирует филогенетическое дерево, основываясь на эволюционной модели, которая обеспечивает максимальное значение функции правдоподобия и с наибольшей вероятностью описывает данные. Выбор эволюционной модели осуществляется с помощью специализированных программ, примерами которых являются ModelEstimator, jModeltest и Prottest (Arvestad, 2006; Darriba *et al.*, 2011; Darriba *et al.*, 2012). Методы правдоподобия получают лучшие результаты, чем методы основанные на расчете расстояний и экономии, но они и менее распространены из-за высоких вычислительных затрат при работе с реальными данными.

Примерами программ, осуществляющих филогенетический анализ последовательностей, являются: PHYLIP (Felsenstein, 1989), PAUP (Swofford, 2000), MRBAYES (Huelsenbeck, Ronquist, 2001) и MEGA (Tamura *et al.*, 2011).

1.7.4. Методы анализа горизонтального переноса МГЭ

ГП является сложным явлением, для определения которого было сформулировано три критерия: неравномерное распределение объекта между видами определенной таксономической группы; высокое сходство последовательностей из геномов репродуктивно изолированных видов, превышающее ожидаемый уровень сходства с учетом времени дивергенции видов хозяев; несоответствие филогении построенной на основе объекта ГП и филогении видов хозяев. Однако, каждый из критериев может являться результатом эволюционных процессов, отличных от ГП, например, стохастической потери и / или дифференциальной фиксации элементов у предковых форм, отбора, действующего на сохранение последовательности МГЭ для выполнения клеточных функций и др. (Shedlock, 2006). Таким образом, в настоящее время не существует единого стандарта и подхода, позволяющего отличить случаи ГП от альтернативных эволюционных процессов.

Методы подтверждения ГП подразделяются на параметрические и филогенетические (Рисунок 1.14). Многие характеристики геномной последовательности являются видоспецифичными («геномная подпись») и позволяют выявить участки, которые были перенесены в геном горизонтально из генома другого вида. *Параметрические методы* позволяют выявить участки генома, отличающиеся от основной последовательности по своим характеристикам: нуклеотидному составу (ГЦ составу и частоте встречаемости синонимичных кодонов (Codon usage bias - CUB)), частоте встречаемости олигонуклеотидов, особенностям «геномного контекста» (окружения участка последовательности, который потенциально был перенесен горизонтально) (Dufraigne *et al.*, 2005; Elhai *et al.*, 2012).

Филогенетические методы анализа подразделяются на явные и неявные. Явные филогенетические методы основаны на сравнении филогенетических деревьев, построенных на основе последовательностей различных генов (МГЭ), с

эталонным (референсным) деревом видов. Данная группа методов обладает рядом недостатков: результат значительно зависит от способа реконструкции филогенетических деревьев; конфликт между филогенетическими деревьями может являться результатом эволюционных событий отличных от горизонтального переноса; количество анализируемых последовательностей ограничивается сложностью подсчетов. В отличие от явных методов неявные основаны на сравнении эволюционных расстояний между последовательностями, не требуют построения филогенетических деревьев, и являются более надежными и экспрессными по сравнению с явными. Неявные методы основаны на гипотезе «молекулярных часов», утверждающей, что эволюционные расстояния между последовательностями генов пропорциональны эволюционным расстояниям между видами (Bromham, Penny, 2003). Основным недостатком неявных методов является зависимость результатов данных методов от лучших хитов BLAST поиска.

Поскольку параметрические и филогенетические методы обладают недостатками и зачастую могут давать противоречивые результаты, для подтверждения горизонтального переноса используется обе группы методов.

Область разработки новых биоинформатических методов изучения МГЭ в настоящее время активно развивается. Получение новых данных о последовательностях МГЭ стимулирует развитие данной области и формирование комплексного подхода к анализу МГЭ. Широкое распространение повторенных последовательностей в геномах осложняет секвенирование, сборку, аннотацию и выравнивание новых геномов. Данный факт также приводит к росту заинтересованности в развитии новых компьютерных методов анализа МГЭ (Bergman, Quesneville, 2007).

Заключение по обзору литературы

Мобильные генетические элементы - разнообразная группа повторенных последовательностей, обладающих способностью менять локализацию в геноме хозяина. МГЭ обнаружены в геномах практически всех эукариотических организмов и относятся к классу умеренно повторенных последовательностей ДНК. МГЭ оказывают значительное влияние на эволюцию геномов хозяев путем изменения структуры и размера генома, предоставления генетического материала

для сборки новых генов и формирования новых функций, обмена генетической информацией между репродуктивно изолированными видами (горизонтальный перенос).

С момента открытия мобильных генетических элементов в 1951 году данные объекты активно изучаются с помощью биоинформатических и экспериментальных методик. Несмотря на актуальность проблемы изучения МГЭ, в картине происхождения, разнообразия и распространения МГЭ остается ряд нерешенных вопросов:

1. Разнообразие, распространение и эволюция МГЭ в геномах эукариот. Исследования в данной области направлены на разработку биоинформатических инструментов для поиска, аннотации и анализа МГЭ в различных группах эукариотических организмов.

2. Роль МГЭ в возникновении новых функций. Не смотря на то, что МГЭ длительное время относились к категории «мусорной ДНК», в настоящее время появляется все больше доказательств того, что они МЭ играют ключевую роль в эволюции геномов эукариот. Основные исследования в данной области направлены на идентификацию и анализ последовательностей (кодирующих и некодирующих), произошедших путем доместикации МГЭ и получивших определенную функцию в геноме хозяина.

3. Механизм горизонтального переноса МГЭ. К настоящему моменту в литературе описано значительное количество случаев ГП МГЭ между геномами эволюционно удаленных видов эукариотических организмов, однако механизм данного процесса практически не изучен. Особый интерес исследователей связан с изучением процесса ГП ДНК транспозонов порядка TIR, для которых ГП является частым событием, и non-LTR ретротранспозонов для которых данное явление встречается исключительно редко.

Глава 2. МАТЕРИАЛЫ И МЕТОДЫ

2.1. Выбор стратегии исследования

Объектами исследования в данной работе являются МГЭ из геномов представителей отряда Lepidoptera. К настоящему моменту секвенирован геном единственного представителя данного отряда – тутового шелкопряда *Bombyx mori*. В базе данных присутствует ограниченная информация о МГЭ других представителей отряда Lepidoptera. Для решения задач, поставленных в рамках данной работы, был применен комплексный подход, предполагающий использование биоинформатических и экспериментальных методик (Рисунок 2.1).

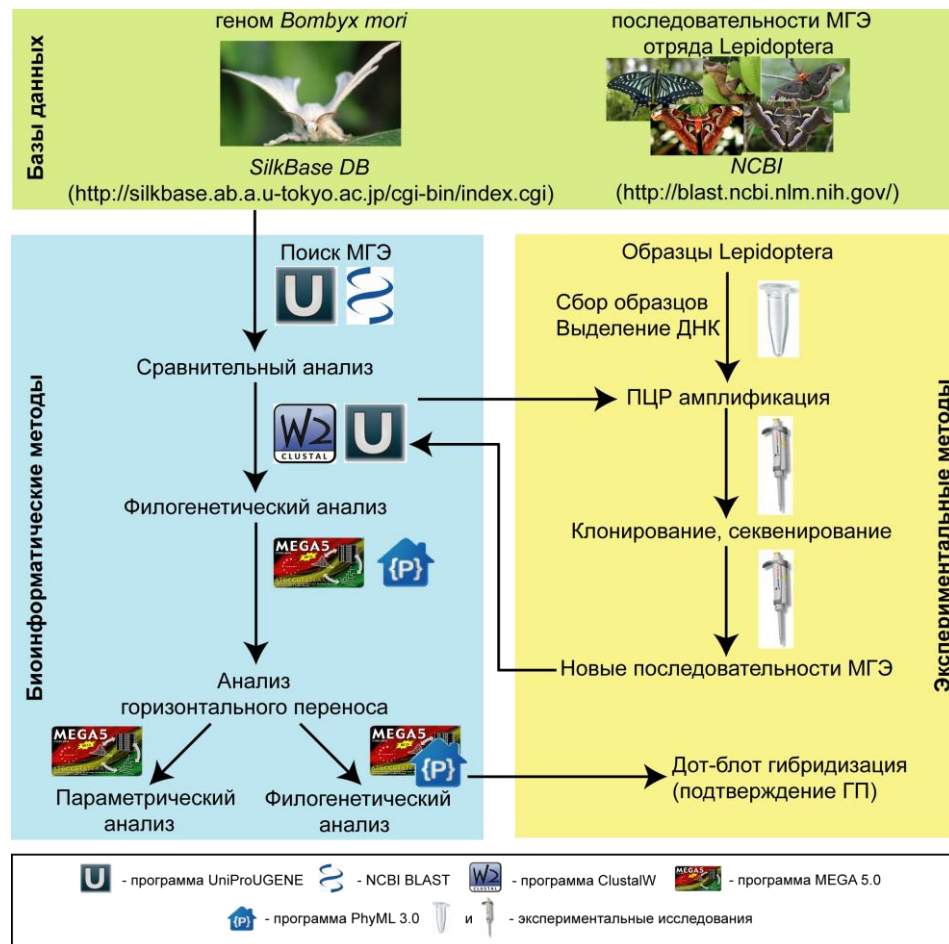


Рисунок 2.1. Схема исследования мобильных генетических элементов (МГЭ) групп non-LTR ретротранспозонов и TIR ДНК транспозонов в геномах представителей отряда Lepidoptera. Биоинформатические этапы исследования представлены в голубом прямоугольнике, экспериментальные этапы в желтом прямоугольнике, базы данных, использованные в исследовании, в зеленом прямоугольнике. В анализе были использованы программы UniPro UGENE (Okonechnikov *et al.*, 2012), MEGA 5.0 (Tamura *et al.*, 2011), PhyML 3.0 (Guindon *et al.*, 2010), ClustalW (Thompson *et al.*, 1994) и он-лайн сервисы NCBI. ГП – горизонтальный перенос.

Основу данного подхода составляют биоинформатические методики, позволяющие проводить поиск, сравнительный и филогенетический анализ последовательностей МГЭ, а также анализ потенциального ГП элементов. Экспериментальные методики в данном случае играют вспомогательную роль и являются поставщиками последовательностей мобильных генетических элементов и геномных последовательностей для новых модельных организмов.

2.2. Биоинформатический анализ

2.2.1. Поиск последовательностей *CR1 non-LTR* ретротранспозонов и *mariner-like* ДНК транспозонов

Поиск по генетическим базам данных. Для поиска МГЭ из геномов представителей отряда Lepidoptera и класса Insecta, для которых отсутствуют полногеномные последовательности, использовалась база данных Genbank (NCBI) (www.ncbi.nlm.nih.gov/). Логический поиск МГЭ в базе данных Genbank проводился по запросам «transposon mariner», «mariner-like» «transposase», «DNA transposon» в случае *mariner-like (MLE)* ДНК транспозонов и «non-LTR retrotransposon», «CR1 element», «retrotransposon», «revertase» в случае *CR1 non-LTR* ретротранспозонов.

Поиск по прочитанным геномам. Последовательность генома тутового шелкопряда *B. mori* представлена в базе данных SilkDB (<http://www.silkdb.org/>, Duan *et al.*, 2010). Для биоинформатического поиска последовательностей МГЭ в геноме *B. mori* использовалось приложение Query Designer пакета программ UniPro UGENE (Okonechnikov *et al.*, 2012). В приложении Query Designer были сконструированы схемы поиска *mariner-like* ДНК транспозонов и *CR1B non-LTR* ретротранспозонов, использующие инструмент HMMER2. Данный инструмент позволяет осуществлять поиск последовательностей с помощью скрытых марковских моделей (HMM - Hidden Markov Model) или HMM профилей, представляющих собой статистическую модель искомой аминокислотной последовательности (Eddy, 1998; <http://hmmer.wustl.edu/>). В случае *MLE* элементов HMM профили были построены на основе выравнивания аминокислотных последовательностей транспозазы, а в случае *CR1B* элементов на основе

выравнивания последовательностей обратной транскриптазы. Последовательности обратной транскриптазы и транспозазы были получены в результате поиска по базе данных GenBank (<http://www.ncbi.nlm.nih.gov>). Последовательности, обнаруженные в результате поиска по геному *B. mori*, были использованы для конструирования олигонуклеотидных праймеров, сравнительного и филогенетического анализа.

Для поиска химерных последовательностей *CR1B/MLE* в схему Query Designer был внесен дополнительный этап, позволяющий анализировать последовательность в направлении 5' и 3' от выявленного элемента. На первом этапе проводился поиск *CR1B* элементов, с использованием схемы для *CR1B* non-LTR ретротранспозонов, затем проводился анализ последовательностей фланкирующих данный элемент с помощью программы Blast (алгоритмы Blastn и tBlastx) (Altschul *et al.*, 1990).

2.2.2. Сравнительный анализ последовательностей МГЭ

Выравнивания нуклеотидных и аминокислотных последовательностей были построены с использованием программы ClustalW (Thompson *et al.*, 1994). Для выравнивания использовались следующие параметры: штраф за открытие разрыва 30, штраф за продолжение разрыва 0.8, весовая матрица IUB. Улучшение выравниваний проводилось с помощью алгоритма MUSCLE доступного в пакете программ UniPRO UGENE v. 1.12.3 (Okonechnikov *et al.*, 2012). Для поиска и анализа открытых рамок считывания использовалась программа ORF finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>).

2.2.3. Филогенетический анализ последовательностей МГЭ

Построение филогенетических деревьев проводили на основе множественных выравниваний последовательностей транспозазы в случае *MLE* ДНК транспозонов, и последовательностей обратной транскриптазы в случае *CR1* non-LTR ретротранспозонов.

Для построения филогенетических деревьев были использованы два метода: метод соединения ближайших соседей (NJ – Neighbor Joining) из группы матричных методов и метод максимального правдоподобия (ML - Maximum Likelihood) из группы символьно-ориентированных методов. NJ филогенетические

деревья были построены при помощи программы MEGA 5.0 (Tamura *et al.*, 2011); ML филогенетические деревья были построены при помощи программы PhyML v.3.0 (Guindon *et al.*, 2010). Программа jModelTest, версии 2.1.4, была использована для выбора наиболее подходящей модели нуклеотидных замен (Darriba *et al.*, 2012). Для оценки достоверности топологии филогенетических деревьев использовался бутстреп-тест (bootstrap test, 1000 репликаций) (Felsenstein, 1985).

2.2.4. Анализа горизонтального переноса МГЭ

Анализ ГП проводился филогенетическими (явными) и параметрическими методами. В качестве параметрических методов использовались «divergence-versus-age» анализ и анализ количества синонимичных замен.

«Divergence-versus-age» анализ позволяет оценить скорость эволюции последовательностей за счет графической визуализации уровня дивергенции последовательностей из различных групп организмов против времени дивергенции организмов-хозяев. Контрольные кривые скорости эволюции последовательностей МГЭ позвоночных и артропод, наследование которых происходило исключительно вертикальным путем, были взяты из литературных данных (Malik *et al.*, 1999). Расположение результатов вблизи контрольных кривых предполагает, что дивергенция последовательностей произошла одновременно с дивергенцией видов хозяев и подтверждает вертикальное наследование. Расположение результатов ниже контрольных кривых предполагает, что дивергенция последовательностей произошла позже дивергенции видов хозяев и подтверждает ГП. Дистанции между аминокислотными последовательностями обратной транскриптазы и транспозазы на сайт, использованные для «divergence-versus-age» анализа, также были подсчитаны с помощью программы MEGA5 (Tamura *et al.*, 2011).

Сравнение числа синонимичных замен на синонимичный сайт (dS) позволяет определить происхождение последовательностей МГЭ. Согласно сценарию вертикального наследования МГЭ dS по последовательности МГЭ между видами должна соответствовать dS консервативных генов между данными видами. Если МГЭ был перенесен в геном горизонтально, то значение dS по последовательности МГЭ должно быть значительно ниже, чем dS консервативных генов. Выравнивания

аминокислотных последовательностей транспозазы использовались для подсчета количества синонимичных замен на сайт (dS) методом Nei, Gojobori (1986) в программе MEGA5 (Tamura *et al.*, 2011). В качестве контроля были использованы консервативные последовательности функциональных генов представителей отряда Lepidoptera: *wg*—*Maculinea alcon* (HQ918074), *M. arion* (HQ918080), *M. nausithous* (HQ918085), *M. teleius* (HQ918090), и *B. mori* (EU033069); *EF1 α* — *M. alcon* (HQ918097), *M. arion* (HQ918114), *M. nausithous* (HQ918116), *M. teleius* (HQ918117), и *B. mori* (EU136667); и *H3*—*M. alcon* (HQ917973), *M. arion* (GQ128823), *M. nausithous* (HQ917976), *M. teleius* (HQ917978), и *B. mori* (DQ443228). Для оценки статистической достоверности разницы между dS различных последовательностей был использован точный тест Фишера.

2.3. Экспериментальные методики

2.3.1. Выделение геномной ДНК

Образцы имаго представителей отряда Lepidoptera были предоставлены доктором Чарльзом Миттером (Мэрилендский университет, США), к.б.н. О. Костериным (Институт цитологии и генетики СО РАН, Россия) и к.б.н. В. Дубатоловым (Институт систематики и экологии животных СО РАН, Россия). Личинки представителей рода *Maculinea* были собраны Евой Сливинской (Ягеллонский университет, Польша). Список видов отряда Lepidoptera использованных в работе представлен в таблице 3.1.

Выделение геномной ДНК проводили из тканей личинок видов рода *Maculinea* и имаго других представителей отряда Lepidoptera, фиксированных в 96% этиловом спирте. Фиксированный материал предварительно высушивали при комнатной температуре в течение 10-15 минут. Выделение тотальной ДНК проводили с использованием набора реактивов DNeasy Blood & Tissue Kit фирмы “QIAGEN” (Германия), согласно протоколу фирмы-производителя.

2.3.2. Полимеразная цепная реакция

Полимеразную цепную реакцию (ПЦР) проводили в пробирках типа «Eppendorf» на амплификаторе «BIS Termocycler» производства фирмы ООО «БИС-Н» в объеме 20 мкл. Реакционная смесь содержала 10 мМ Трис-НСl (рН 8.4), 50 мМ КСl, 1.5 мМ MgCl₂, 0.2 мМ dTTP, 0.2 мМ dGTP, 0.2 мМ dATP, 0.2 мМ dCTP, 20 мкМ каждого олигонуклеотидного праймера, 1 единиц активности (е.а.) Taq ДНК полимеразы («Invitrogen», США), 0.1 нг геномной ДНК. Программа ПЦР амплификации включала в себя предварительную денатурацию (3 мин. при 94°C) и 30 циклов, включающих: денатурацию 30 сек. при 94°C, отжиг праймеров 30 сек. при 42-56 °С (в зависимости от характеристики олигонуклеотидных праймеров) и элонгацию 1 - 2 мин. при 72°C (в зависимости от длины фрагмента).

Перечень олигонуклеотидных праймеров, использованных в работе, приведен в Приложении 1. Расчет олигонуклеотидных праймеров для ПЦР амплификации элемента *BmmarY* и химерных *CR1B/MLE* проводился с использованием программы VectorNTI v.11.0 (Lu, Moriyama, 2004).

2.3.3. Электрофорез продуктов ПЦР в агарозном геле

Разделение продуктов ПЦР реакций проводили с помощью электрофореза в 1% агарозном геле, приготовленном на буфере ТАЕ состава: 40 мМ Трис-НСl (рН 8.0), 20 мМ NaAc, 1 мМ ЭДТА и 0.01 мкг/мл бромистого этидия. Анализируемые образцы ДНК смешивали с буфером для нанесения образцов, содержащим 50% глицерина, 0.3% бромфенолового синего и 0.3% ксиленцианола, в соотношении 1:10. Электрофорез проводили в течение 15-20 мин. при напряженности электрического поля 3-7 В/см в камере для горизонтального электрофореза Wide Mini Sub Cell GT («Bio-Rad», США). Визуализацию фрагментов ДНК осуществляли в ультрафиолетовом свете (254 нм). Электрофореграмму документировали при помощи трансиллюминатора ETX-20.C (Vilber Lourmat) и системы гель-документирования Gel Imager-2 (Хеликон, Россия).

Элюцию целевого фрагмента ДНК из геля проводили с использованием набора QIAquick фирмы «Qiagen» (Германия), согласно протоколу фирмы-производителя.

2.3.4. Клонирование фрагментов ДНК

2.3.4.1. Лигирование фрагментов с T-вектором

Продукты ПЦР были лигированы с T-вектором pGEM-T Easy (Promega, США). Реакционная смесь содержала 0.15 мкг. амплифицированного фрагмента ДНК, 0.1 мкг. T-вектора, 5 е.а. T4 ДНК-лигазы («Invitrogen», США), 10 mM Трис-НСl (рН 8.5, 37°C), 10 mM MgCl₂, 100 mM KCl и 1 mM АТР. Реакцию проводили при температуре 22 °С в течении 1-2 часов. Продукты лигирования были использованы для трансформации химически компетентных клеток *E. coli* штамма XL10-Gold.

2.3.4.2. Трансформация компетентных клеток *E.coli*

Компетентные клетки размораживали на льду, добавляли лигазную смесь, выдерживали на льду в течение 30 мин. и инкубировали в течение 2 мин. при 42°C (тепловой шок). К клеточной суспензии добавляли 1 мл среды Лурия-Бертани (LB: 1% NaCl, 1% бакто-триптон, 0.5% бакто-дрожжевой экстракт, рН 7.5), и инкубировали в течение 2 часов при 37°C. Далее клеточную суспензию высевали на селективную среду LB, содержащую 25-50 мкг/мл ампицилина, 0.002% X-Gal и 100 μМ изопропил-β-D-1-тиогалактопиранозид (IPTG). Инкубировали в термостате при 37°C в течение 12-16 часов, после чего хранили при 4°C.

2.3.5. Отбор и анализ клонов-трансформантов

Для отбора клонов-трансформантов, содержащих плазмиду со встройкой целевого фрагмента ДНК, использовали устойчивость трансформированных бактерий к ампицилину и отсутствие способности к α-комплементации (белоголубая селекция). Процедура анализа отобранных клонов включала ПЦР амплификацию и секвенирование. Для ПЦР амплификации использовались праймеры M13, комплементарные участкам плазмиды, фланкирующим целевой фрагмент (M13-reverse: 5'-GGAAACAGCTATGACCATG-3', M13-forward: 5'-GTAAAACGACGGCCAGT-3').

Реакции ПЦР амплификации проводили на амплификаторе «BIS Thermocycler» (ООО «БИС-Н») в объеме 20 мкл. Реакционная смесь содержала 10 mM Трис-НСl (рН 8.4), 50 mM KCl, 1.5 mM MgCl₂, 0.2 mM dTTP, 0.2 mM dGTP, 0.2

мМ dATP, 0.2 мМ dCTP, 20 мкМ каждого олигонуклеотидного праймера, 2.5 единиц активности (е.а.) Taq ДНК полимеразы («Invitrogen», США). Программа ПЦР амплификации включала в себя предварительную денатурацию (3 мин. при 94°C) и 30 циклов, включающих: денатурацию 30 сек. при 94°C, отжиг праймеров 30 сек. при 52 °С и элонгацию 1-2 мин. при 72°C (в зависимости от длины фрагмента).

Продукты реакции анализировали электрофорезом в 1% агарозном геле. Продукты ожидаемой длины выделяли из геля и использовали для определения нуклеотидной последовательности.

2.3.6. Установление нуклеотидных последовательностей

Установление нуклеотидных последовательностей проводили методом автоматического секвенирования. Включение метки проводилось при помощи реакции Сэнгера с использованием реагента BigDye Terminator v. 3.1 (Applied Biosystems) (Sanger *et al.*, 1977). Циклическая реакция секвенирования проводилась в объеме 10 мкл и включала предварительную денатурация (94°C, 3 мин.) и 20 циклов, включающих денатурацию (96°C, 30 сек.), отжиг (50°C, 30 сек.) и элонгацию (60°C, 4 мин.). Реакционная смесь содержала 2 мкл буфера BigDye Terminator 5X Sequencing Buffer, 2мкл BigDye Terminator v. 3.1, 50-100 нг ДНК-матрицы и 0.32 мкМ праймера.

Продукт реакции осаждали изопропанолом: на 10 мкл реакционной смеси добавляли 30 мкл воды и 60 мкл 100% изопропанола и тщательно перемешивали. Оставляли при комнатной температуре на 30 мин., после чего центрифугировали 10 мин. при 13200 об./мин. Супернатант отбирали и добавляли 100 мкл 70% этилового спирта, снова центрифугировали 5 мин. при 13200 об./мин. Аккуратно отбирали супернатант и сушили осадок при комнатной температуре.

Определение нуклеотидной последовательности проводилось в центре коллективного пользования «Геномика» СО РАН. Полученные последовательности ДНК транспозонов, non-LTR ретротранспозонов и химерные последовательности *CR1B/MLE* были выложены в базу данных GenBank (номера карточек: JQ580972-JQ580974, HQ284259-HQ284570, HQ606330-HQ606381, HQ606387-HQ606390, HQ606392-HQ606396, HQ606399, HQ606401, HQ606402, HQ606405-HQ606462).

2.3.7. Дот-блот гибридизация и создание библиотеки клонов

Дот-блот гибридизация осуществлялась следующим образом: геномная ДНК представителей отряда Lepodoptera в количестве 500 нг и 100 нг наносилась на нейлоновые мембраны Nytran SuPer Charge согласно протоколу фирмы производителя и обрабатывалась УФ с помощью Bio-Rad GS Genelinker (Bio-Rad, США). Далее проводили гибридизацию в течении 1 часа при 65°C в растворе QuikHyb Hybridization Solution (Stratagene), с добавлением тотальной ДНК спермы лосося в концентрации 100 г/мл и ДНК пробами меченными по ^{32}P с помощью набора реактивов DNA Polymerase I Large (Klenow) Fragment Mini Kit (Promega, США). Мембрану отмывали дважды в промывочном буфере по 45 минут при температуре 65°C. Сигналы гибридизации детектировали с помощью экспозиции на рентгеновскую пленку с течение 12-16 часов.

Для детекции *BmmarY*-like *MLEs* в качестве зонда был использован P^{32} -меченный фрагмент последовательности *MtemarY1* элемента из *M. teleius*, для детекции *Bmmar1*-like элементов был использован P^{32} -меченный фрагмент последовательности *Mtemar1.2* из *M. arion*. Для приготовления пробы на *CR1B* элементы, 5'-UTR элемента *CR1B* из генома *M. teleius* был амплифицирован с помощью специфической пары праймеров (Приложение 1).

Мини библиотека была создана для четырех видов: *M. teleius*, *M. arion*, *B. mori*, и *Bombux mandarina*. Для амплификации химерных конструкторов *CR1B/MLE* была использована ПЦР амплификация с праймерами на химерные элементы, перечисленными в Приложении 1. Полученные фрагменты были разделены с помощью электрофореза в агарозном геле и выделены с помощью набора реагентов QIAquick Gel Extraction Kit (QIAGEN). Очищенные фрагменты были лигированы с вектором pGEM с использованием набора реактивов pGEM-T cloning kit (Promega). Анализ присутствия *CR1B* и *MLE* элементов в клонах использовался «colony lift assay» согласно Sambrook *et al.* (1989). Для гибридизации ^{32}P проб с минибibliothekой клонов использовался протокол аналогичный вышеописанному для дот-блот гибридизации. На первом этапе проводилась гибридизация с пробами специфичными к *MLEs* (*Bmmar1* или *BmmarY*). Затем данная мембрана вновь была использована для гибридизации с пробами специфичными к *CR1B* элементам.

ГЛАВА 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1. *CR1* non-LTR ретротранспозоны в геномах представителей отряда *Lepidoptera*

Объектами исследования в настоящей работе являются МГЭ из геномов представителей отряда *Lepidoptera* (Чешуекрылые). Отряд *Lepidoptera* – самый крупный отряд насекомых (*Insecta*), включающий в себя более 157000 видов (van Nieukerken *et al.*, 2011). Согласно современной классификации в отряде выделяют 47 суперсемейств, 124 семейства и 332 подсемейства (Kristensen, 2003). Основное внимание в данной работе будет уделено МГЭ из геномов филогенетической группы *Macrolepidoptera*. Филогения данной группы представлена на рисунке 3.1.

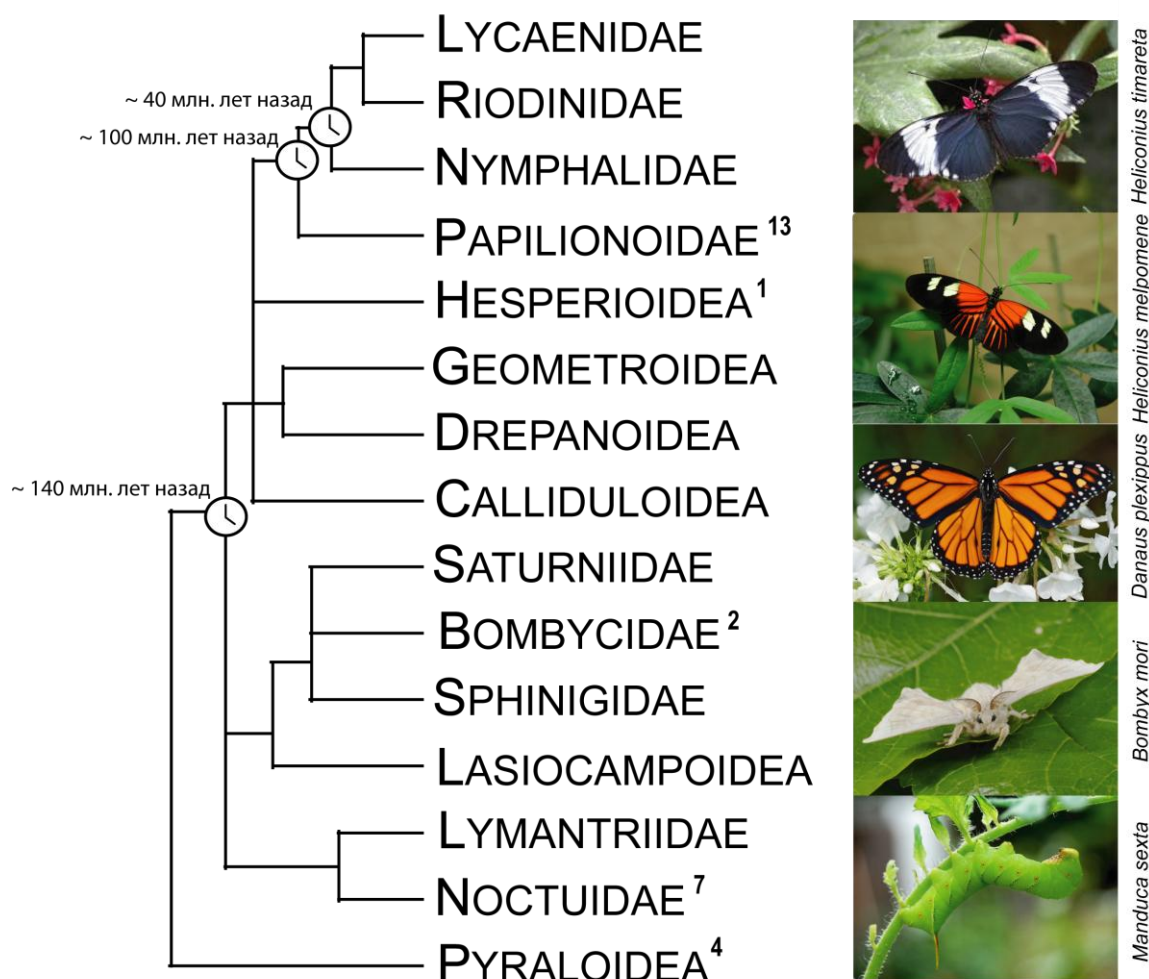


Рисунок 3.1. Филогенетическое древо суперсемейств филогенетической группы *Macrolepidoptera*, использованных в работе (Tree of Life Web Project (<http://tolweb.org/>) с дополнениями). Цифры обозначают количество геномных проектов в каждом суперсемействе. В правой части рисунка приведены изображения представителей *Macrolepidoptera*, для которых геномные проекты находятся на завершающей стадии.

Представители отряда Lepidoptera являются значимыми компонентами природных экосистем, оказывают влияние на деятельность человека и являются удобными модельными объектами для научных исследований. В настоящее время отряд Lepidoptera представляет большой интерес для биологов различных специальностей. На сегодняшний день полностью секвенирован геном тутового шелкопряда *Bombyx mori* (суперсемейство Bombycoidea) и на завершающей стадии находятся геномные проекты четырех представителей Macrolepidoptera: *Heliconius melpomene* (суперсемейство Papilionoidea), *Heliconius timareta* (суперсемейство Papilionoidea), *Danaus plexippus* (суперсемейство Papilionoidea) и *Manduca sexta* (суперсемейство Bombycoidea). В процессе развития идут проекты по секвенированию 22 геномов представителей Macrolepidoptera.

Активные исследования проводят в области изучения разнообразия и распространения МГЭ из геномов представителей отряда Lepidoptera. В недавних исследованиях был выявлен уникальный случай горизонтального переноса *CRIB* non-LTR ретротранспозонов между представителями семейств Bombycidae и Lycaenidae (Novikova *et al.*, 2007). Настоящая работа направлена на изучение распространения и разнообразия *CRIB* non-LTR ретротранспозонов и *mariner*-like (*MLE*) TIR ДНК транспозонов в геномах представителей отряда Lepidoptera и выявление механизмов горизонтального переноса.

3.1.1. Поиск *CR1* non-LTR ретротранспозонов в геномах представителей отряда Lepidoptera

Для экспериментального поиска *CR1* non-LTR ретротранспозонов в геномах представителей отряда Lepidoptera были выбраны 60 видов, представляющие 15 подсемейств из 8 суперсемейств (Таблица 3.1). Наиболее широко в выборке представлены виды семейств Bombycidae и Lycaenidae, для которых ранее был выявлен горизонтальный перенос *CRIB* non-LTR ретротранспозонов (Novikova *et al.*, 2007).

Поиск проводился методом ПЦР амплификации с использованием дегенеративных праймеров специфических к последовательностям обратной транскриптазы *CR1* элементов (Приложение 1). Положительный сигнал ПЦР амплификации был получен для 55 из 60 исследованных видов (Таблица 3.1).

Таблица 3.1. Список исследованных видов отряда Lepidoptera, их таксономический статус и результаты ПЦР анализа с праймерами, специфичными к *CRI* и *MLE* элементам.

Суперсемейство	Семейство	Вид	ПЦР <i>CRI</i>	ПЦР <i>MLE</i> **
Papilionoidea	Lycaenidae	<i>Plebejus argus</i>	+	-
		<i>Scolitantides orion</i>	+	+
		<i>Shijimaoides divina</i>	+	+
		<i>Pseudozizeeria maha</i>	-	+
		<i>Maculinea teleius</i>	+	+(1)
		<i>Maculinea arion</i>	+	+(2)
		<i>Maculinea nausithous</i>	+	+
		<i>Maculinea alcon</i>	+	+
		<i>Pseudolucia collina</i>	+	+
		<i>Thecla betulae</i>	+	-
		<i>Narathura japonica</i>	+	+
		<i>Brangas neora</i>	+	+
		<i>Eumaeus godarti</i>	+	+(3)
		<i>Curetis regula</i>	-	+(4)
	Riodinidae	<i>Anteros formosus</i>	+	+
		<i>Emesis lucinda</i>	+	+
		<i>Euselasia chrysippe</i>	+	+
		<i>Catocyclotis adelina</i>	+	+

* - выделены виды, для которых по результатам секвенирования не были выявлены *CRI* элементы.

** - цифрами выделены виды, использованные для дот-блот гибридизации.

Таблица 3.1 (продолжение). Список исследованных видов отряда Lepidoptera, их таксономический статус и результаты ПЦР анализа с праймерами, специфичными к *CRI* и *MLE* элементам.

Суперсемейство	Семейство	Вид	ПЦР <i>CRI</i>	ПЦР <i>MLE</i> **
Papilionoidea	Riodinidae	<i>Theope virgilius</i>	+	+
		<i>Symmachia xypete</i>	+	-
		<i>Mesosemia lamachus</i>	+*	- (5)
		<i>Caria rhacotis</i>	+	+
	Nymphalidae	<i>Araschnia levana</i>	+	+(6)
		<i>Melitaea phoebe</i>	+	+(7)
		<i>Erebia thean</i>	+	+(8)
		<i>Coenonympha glycerion</i>	+	-
		<i>Oeneis magna dubia</i>	+*	-
		<i>Oeneis sculda</i>	+*	+(9)
	Papilionidae	<i>Parnassius stubbendortii</i>	+	- (10)
	Pieridae	<i>Colias hyale</i>	-	- (11)
Hesperioidea	Hesperiidae	<i>Heteropterus morphaeus</i>	+	-
Geometroidea	Geometridae	<i>Scopula ornata</i>	+	-
		<i>Semiothisa clathrata</i>	+	-
		<i>Calospilos sylvata</i>	+	-
		<i>Brephos parthenius</i>	+*	-

* - выделены виды, для которых по результатам секвенирования не были выявлены *CRI* элементы.

** - цифрами выделены виды, использованные для дот-блот гибридизации.

Таблица 3.1 (продолжение). Список исследованных видов отряда Lepidoptera, их таксономический статус и результаты ПЦР анализа с праймерами, специфичными к *CRI* и *MLE* элементам.

Суперсемейство	Семейство	Вид	ПЦР <i>CRI</i>	ПЦР <i>MLE</i> **
Drepanoidea	Drepanidae	<i>Drepana sp.</i>	+	+(12)
Calliduloidea	Callidulidae	<i>Callidula sp.</i>	+*	+
Noctuoidea	Lymantriidae	<i>Lymantria dispar</i>	+	+(13)
	Noctuidae	<i>Polia nebulosa</i>	+	+(14)
		<i>Mythimna sp.</i>	+	+(15)
		<i>Agrotis exclamationis</i>	+	+
Bombycoidea	Sphinigidae	<i>Erinnyis ello</i>	+	-
		<i>Perigonia ilus</i>	+	-
		<i>Agrius cingulata</i>	+*	-
		<i>Dovania poecila</i>	+	-
		<i>Paonias myops</i>	-	-(17)
	Saturniidae	<i>Aglia tau</i>	+*	-(18)
		<i>Janiodes laverna</i> <i>nigropunkta</i>	+*	-
		<i>Asthenidia transversaria</i>	+	-
		<i>Oxytenis modestia</i>	+	-(19)
	Bombycidae	<i>Bombyx huttoni</i>	+	+

* - выделены виды, для которых по результатам секвенирования не были выявлены *CRI* элементы.

** - цифрами выделены виды, использованные для дот-блот гибридизации.

Таблица 3.1 (продолжение). Список исследованных видов отряда Lepidoptera, их таксономический статус и результаты ПЦР анализа с праймерами, специфичными к *CRI* и *MLE* элементам.

Суперсемейство	Семейство	Вид	ПЦР <i>CRI</i>	ПЦР <i>MLE</i> **
Bombycoidea	Bombycidae	<i>Bombyx mori</i>	+	+(20)
		<i>Bombyx mandarina</i>	+	+(21)
		<i>Colla glaucescens</i>	-	-(22)
		<i>Oberthueria caeca</i>	+	+
		<i>Ernolatia moorei</i>	+	+
		<i>Quentalia chromana</i>	+*	+(23)
		<i>Trilocha sp.</i>	+*	-
		<i>Truncina brunnea</i>	+*	+(24)
Lasiocampoidea	Lasiocampidae	<i>Lasiocampa quercus</i>	+	+(25)

* - выделены виды, для которых по результатам секвенирования не были выявлены *CRI* элементы.

** - цифрами выделены виды, использованные для дот-блот гибридизации.

ПЦР продукты были клонированы в плазмидном векторе и для всех клонов были установлены нуклеотидные последовательности. Полученные последовательности были проанализированы на наличие фрагмента обратной транскриптазы с помощью программы Blast (Altschul *et al.*, 1990). Последовательности 318 клонов показали сходство с обратной транскриптазой non-LTR ретротранспозонов. Последующий сравнительный анализ показал, что основная масса последовательностей (156 клонов из 318) содержала фрагмент обратной транскриптазы, близкой к обратной транскриптазе *CRI* non-LTR ретротранспозонов. Оставшаяся часть последовательностей принадлежала элементам из филогенетических групп *TIQ*, *Jockey* и *RI* non-LTR ретротранспозонов. В общей сложности последовательности *CRI* non-LTR

ретротранспозонов были выявлены в геномах 37 из 60 проанализированных видов. Возможной причиной негативного результата поиска *CR1* non-LTR ретротранспозонов в геномах 23 видов является полная элиминация данных элементов из геномов данных видов. Высокая дивергенция последовательностей в области комплементарной праймерам также может объяснять негативные результаты ПЦР амплификации. Названия элементов были образованы следующим образом: *ScoOriT1Q-38* - *Scolitantides orion* элемент ***T1Q*** клон **38**. Последовательности были выложены в базу данных GenBank под номерами: HQ284259–HQ284570.

3.1.2. Филогенетический анализ *CR1* non-LTR ретротранспозонов из геномов представителей отряда *Lepidoptera*

Для филогенетического анализа были использованы последовательности, полученные в настоящей работе, а также последовательности доступные из базы данных GenBank: *BmCR1B* (Novikova *et al.*, 2007) и *Kendo* (GenBank: AB126052) из генома *B. mori*; *CR1A* и *CR1B* элементов из геномов бабочек рода *Maculinea* (Novikova *et al.*, 2007); последовательности ретротранспозонов из генома *Spodoptera frugiperda* (Noctuidae; GenBank: FP340404 и FP340409) (Рисунок 3.2).

На филогенетическом древе можно выделить пять крупных кластеров: *T1Q*, *Jockey*, *R1*, *Aurivillius* и *Fabre* (Рисунок 3.2). Элементы клада *CR1* non-LTR ретротранспозонов сформировали два кластера *Aurivillius* и *Fabre*. Элементы *Fabre* широко представлены в геномах представителей отряда *Lepidoptera* и были выявлены для 20 видов из 11 семейств отряда, в том числе в геномах настоящих бабочек (семейств *Lycaenidae*, *Riodinidae*, *Nymphalidae* и *Papilionidae*) и молей (семейства *Bombycidae*). Элементы кластера *Aurivillius* характеризуются более ограниченным распространением в геномах *Lepidoptera* и представлены в 14 видах из шести семейств: *Lasiocampidae*, *Saturniidae*, *Nymphalidae*, *Geometridae*, *Sphingidae* и *Noctuidae*.

Жизненный цикл non-LTR ретротранспозонов в геномах эукариот, как правило, включает: внедрение в геном, образование новых копий элементов, активную ретротранспозицию, снижение активности транспозиции, накопление

мутаций, элиминацию из геномов одних групп, ГП и сохранение в геномах других групп организмов.

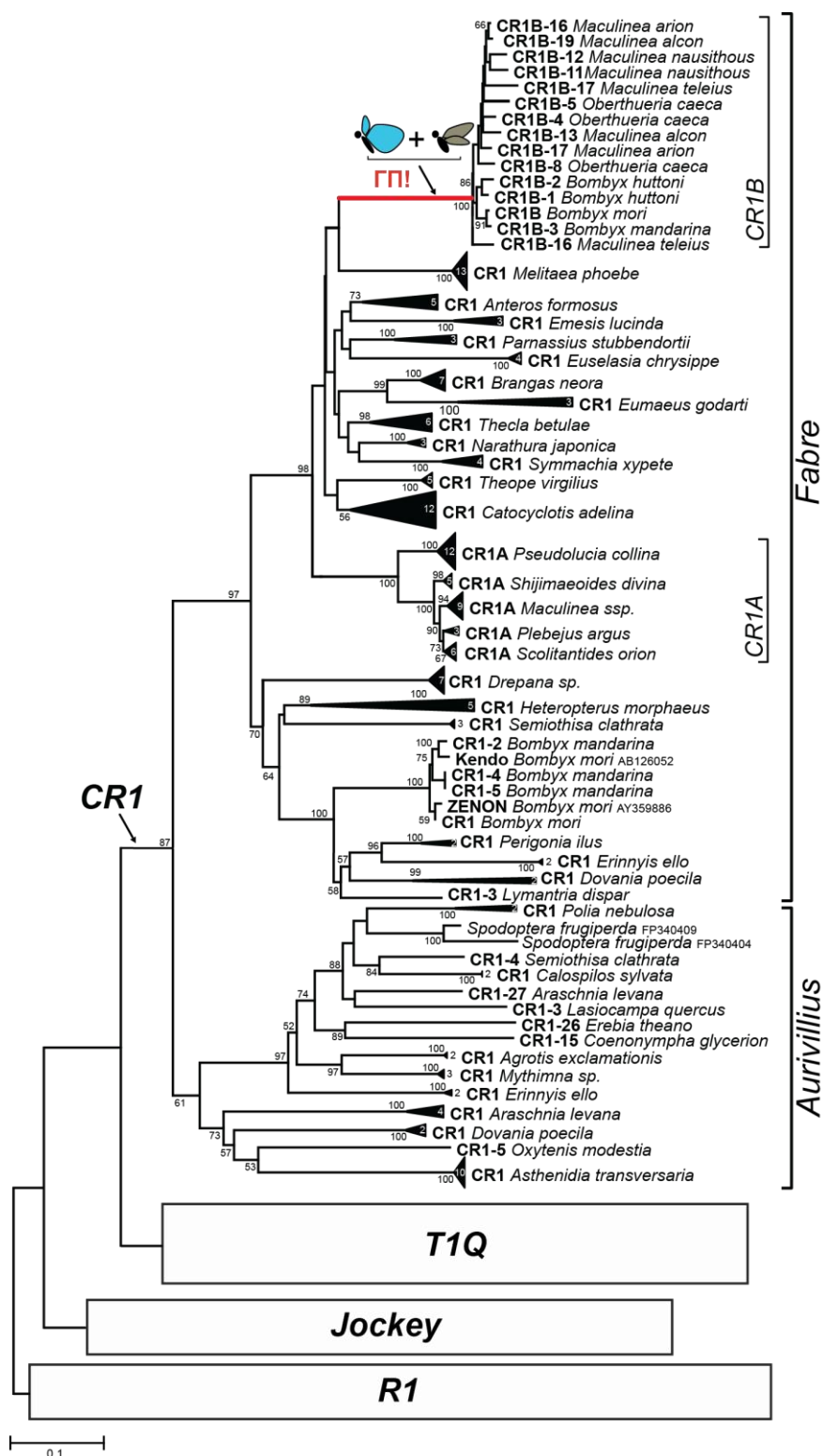


Рисунок 3.2. Бескорневое филогенетическое древо, реконструированное на основе последовательности обратной транскриптазы методом объединения ближайших соседей (NJ) в программе MEGA 5.0. Для оценки достоверности был использован бутстреп-тест (1000 репликаций). Значения коэффициентов поддержки бутстреп-теста менее 50% не показаны. Кластеры *T1Q*, *Jockey* и *R1* представлены схематично. Бабочка голубого цвета обозначает бабочек голубянок семейства *Lycaenidae*, бабочка коричневого цвета – молей семейства *Bombycidae*.

Некоторые из перечисленных процессов могут приводить к несоответствию филогении видов хозяев, построенной на основе последовательностей non-LTR ретротранспозонов и филогении, построенной на основе последовательностей консервативных ядерных генов. Однако, филогения, реконструированная на основе *CR1* non-LTR ретротранспозонов, в основном согласуется с филогенией видов хозяев (Рисунок 3.1, рисунок 3.2). Данный факт свидетельствует о том, что для *CR1* элементов из геномов *Lepidoptera* предпочтительным является вертикальное наследование и сохранение в геноме копий, характеризующихся низким разнообразием.

Несоответствие филогении наблюдается только в случае *CR1*-like элементов кластера *Fabre* (Рисунок 3.2). Так, в подгруппе *CR1B* присутствуют элементы исключительно из геномов семейств, *Bombycidae* и *Lycaenidae*, которые кластеризуются в одну ветвь. Другие подгруппы кластера *Fabre*, напротив, содержат последовательности из различных видов и родов. Сравнительный анализ последовательностей *CR1B* показал, что данные элементы характеризуются высоким сходством последовательностей, как при внутригрупповых сравнениях, так и при сравнениях между семействами *Bombycidae* и *Lycaenidae*. Например, различие *CR1B* элементов на уровне нуклеотидных последовательностей внутри семейств составляет от 1.8% до 4%. Различие между последовательностями из геномов *Bombycidae* и *Lycaenidae* составляет от 3.6% до 4.8%. Согласно литературным данным различие консервативных геномных последовательностей EF-1 α между данными родам составляет 7% (Novikova *et al.*, 2007).

3.1.3. Распространение *CR1* элементов в геномах представителей отряда *Lepidoptera*

Согласно результатам ПЦР амплификации и филогенетического анализа, элементы группы *CR1B* характеризуются неравномерным распределением в геномах изученных видов отряда *Lepidoptera*. Для подтверждения результатов филогенетического анализа была проведена дот-блот гибридизация. Анализ проводился на геномной ДНК 25 видов отряда *Lepidoptera*, в том числе *M. teleius*, *M. arion*, *B. mori* и *B. mandarina* (Рисунок 3.3, таблица 3.1).

Достоверные сигналы гибридизации с пробой, специфичной к элементу *CR1B*, были выявлены только для образцов ДНК представителей родов *Bombyx* и *Maculinea*. Таким образом, результаты дот-блот гибридизации подтверждают результаты ПЦР амплификации и ограниченное распространение *CR1B* элементов в геномах изученных видов отряда *Lepidoptera*.

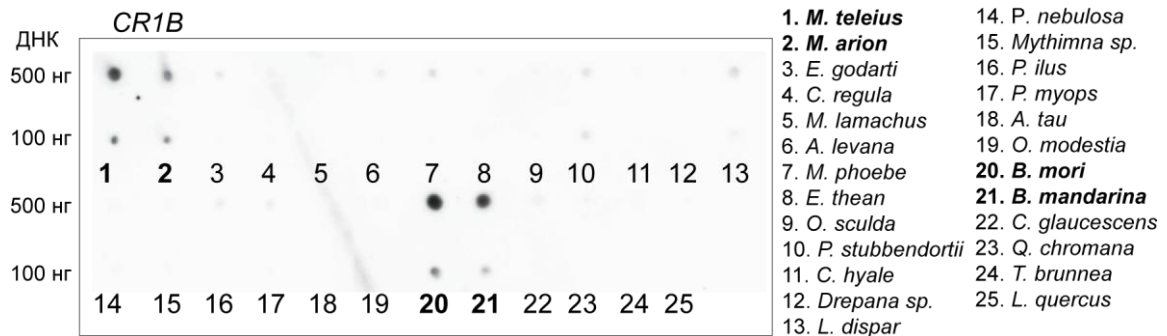


Рисунок 3.3. Результат дот-блот гибридизации образцов ДНК 25 видов бабочек и молей с пробами специфичными к элементу *CR1B*. Список использованных видов приведен справа, номера соответствуют номерам в таблице 3.1. В качестве зонда был использован P^{32} -меченный фрагмент последовательности обратной транскриптазы *CR1B* элемента из *M. teleius*.

3.1.4. Вертикальная эволюция и горизонтальный перенос *CR1* элементов в геномах представителей отряда *Lepidoptera*

Несмотря на то, что, как правило, эволюция non-LTR ретротранспозонов происходит вертикально, в редких случаях выявляется ГП данных элементов. Примерами горизонтального переноса non-LTR ретротранспозонов являются: перенос *Bov-B* элементов между жвачными животными и змеями (Kordis, Gubensek, 1998; Zupunski *et al.*, 2001); перенос элементов между близкородственными видами рода *Drosophila* (Sanchez-Gracia *et al.*, 2005), перенос *Tad* элементов между грибами родов *Eurotiomycetes* и *Sordariomycetes* (Novikova *et al.*, 2009). В рамках данной работы интерес представляет случай потенциального ГП non-LTR ретротранспозонов группы *CR1B*, описанный при изучении представителей рода *Maculinea* (*Lepidoptera*). Ранее также было показано высокое сходство последовательностей *CR1B* элементов из геномов *Maculinea* (*MacCR1B*) и *Bombyx* (*BmCR1B*) (Novikova *et al.*, 2007). В настоящей работе было продемонстрировано неравномерное распространение *CR1B* элементов среди представителей *Macrolepidoptera*, и потенциальный случай ГП был подтвержден с помощью филогенетических методов анализа.

В дальнейшем для детального подтверждения ГП *CR1B* non-LTR ретротранспозонов использовался параметрический метод анализа: «divergence-versus-age» анализ. В рамках данного метода проводится графическая визуализация скорости эволюции последовательностей: уровня дивергенции последовательностей из различных групп организмов против времени дивергенции организмов-хозяев. Метод «divergence-versus-age» активно применяется для анализа горизонтального переноса non-LTR ретротранспозонов (Kordis, Gubensek, 1998; Malik *et al.*, 1999; Novikova *et al.*, 2007; Novikova *et al.*, 2009).

Для сравнения были использованы следующие временные точки:

- 1) дивергенция отрядов Diptera и Lepidoptera ~ 340–380 Mya (million years ago - миллионы лет назад) (Douzery *et al.*, 2004);
- 2) дивергенция суперсемейств Macrolepidoptera (в том числе Bombycoidea и Papilionoidea) ~140 Mya (Gaunt, Miles, 2002);
- 3) дивергенция семейств Lycaenidae и Papilionidae ~ 100 Mya и дивергенция семейств Lycaenidae и Nymphalidae ~ 43 Mya (Nazari *et al.*, 2007).

Значение дистанции для аминокислотных последовательностей обратной транскриптазы *R1*, *Jockey*, *TIQ* и *CR1* элементов, подсчитанные в программе MEGA 5.0, были изображены на графике против времени дивергенции видов хозяев (Рисунок 3.4). В качестве контроля служат кривые скорости эволюции последовательностей МГЭ позвоночных и артропод, наследование которых происходило исключительно вертикальным путем (Malik *et al.*, 1999).

Значения дистанции для элементов *R1*, *Jockey* и *TIQ* из геномов представителей отряда Lepidoptera располагаются вблизи кривой артропод, что свидетельствует об отсутствии событий горизонтального переноса в эволюции данных элементов в геномах изученных видов. По-видимому, последовательности элементов *R1*, *Jockey* и *TIQ* дивергировали в тоже время, когда произошла дивергенция видов хозяев. Аналогичные точки для *CR1* элементов располагаются значительно ниже контрольных кривых (Рисунок 3.4). Таким образом, маловероятно, что дивергенция последовательностей *CR1* элементов произошла одновременно с дивергенцией видов хозяев.

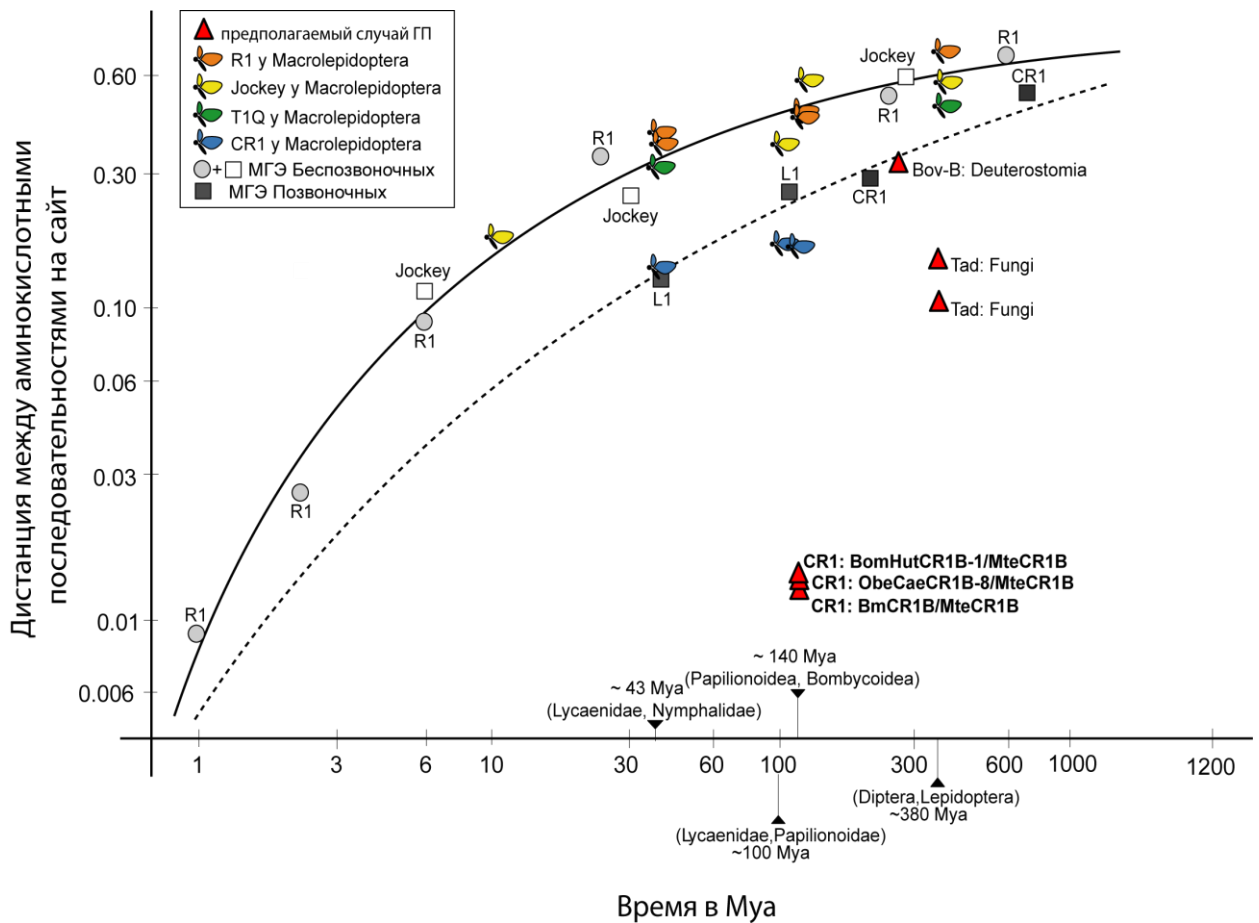


Рисунок 3.4. Результаты «divergence-versus-age» анализа non-LTR ретротранспозонов из геномов Lepidoptera. По оси Y отложены значения различий между аминокислотными последовательностями обратной транскриптазы из различных групп организмов, подсчитанные в программе MEGA. По оси X отложено время дивергенции организмов-хозяев в Mya (Mya - million years ago - миллионы лет назад) (Gaunt, Miles, 2002; Douzery *et al.*, 2004; Nazari *et al.*, 2007). Штрихпунктирная и сплошная линии представляют собой кривые скорости эволюции последовательностей МГЭ позвоночных и артропод, соответственно (согласно Malik *et al.*, 1999).

Наиболее ярко снижение скорости эволюции наблюдается в случае *BomHutCR1B-1* из генома *B. huttoni* (Bombycidae) против *MteCR1B* из *M. teleius* (Lycaenidae); *ObecaeCR1B-8* из *O. caeca* (Bombycidae) против *MteCR1B*, и *BmCR1* из *B. mori* (Bombycidae) против *MteCR1B* из *M. teleius* (временная точка 140 Ma) (Рисунок 3.4).

Снижение скорости эволюции подтверждает ГП данных элементов между Lycaenidae и Bombycidae. Единственным альтернативным объяснением снижения скорости эволюции последовательности является высокое давление отбора, в связи с высокой значимостью функции последовательности для организма-хозяина.

Однако, данный сценарий не может объяснять снижение скорости эволюции у *CR1B* элементов.

Таким образом, вышеописанные факты подтверждают ГП *CR1B* non-LTR ретротранспозонов между бабочками *Lycaenidae* и молями *Bombycidae*.

Механизм ГП non-LTR ретротранспозонов не известен. Согласно одной из гипотез в ГП non-LTR ретротранспозонов принимают участие ДНК транспозоны. Non-LTR ретротранспозон может встроиться в последовательность ДНК транспозона и переместиться в геном нового вида в виде химерного конструкта. Первым подтверждение возможности данного сценария является обнаружение встройки non-LTR ретротранспозона *HpaI* в последовательность *Tc1-like* ДНК транспозона в геномах лососевых (Takasaki *et al.*, 1996). В настоящей работе было сделано предположение о том, что *mariner-like* TIR ДНК транспозоны (*MLEs* – *mariner-like elements*) принимают участие в ГП *CR1B* non-LTR ретротранспозонов между представителями отряда *Lepidoptera*.

3.1.5. Поиск химерных *CR1B/MLE* элементов в геноме *B. mori*

Поскольку ГП произошел относительно недавно в геномах донора и реципиента должны были сохраниться химерные конструкции, принимавшие участие в ГП. Биоинформатический поиск химерных последовательностей типа «ДНК транспозон/*CR1B* элемент» в геноме *B. mori* проводился с помощью приложений Query Designer и HMM пакета программ UGENE (Eddy, 1998; Okonechnikov *et al.*, 2012). Для поиска *CR1B* элементов был сконструирован HMMER2 профиль на основе выравнивания аминокислотных последовательностей обратной транскриптазы, полученных в рамках данной работы.

В результате проведенного поиска были получены HMM сигналы последовательностей *CR1B*, данные последовательности были проанализированы и также были проанализированы геномные районы вблизи данных элементов с помощью различных алгоритмов программы Blast. В результате поиска было обнаружено два локуса (*nscaf2823* и *nscaf2901*), содержащих *Bmmar MLE* ДНК транспозоны (*Bmmar1* и *Bmmar6*) и *BmCR1B* элемент (Рисунок 3.5).

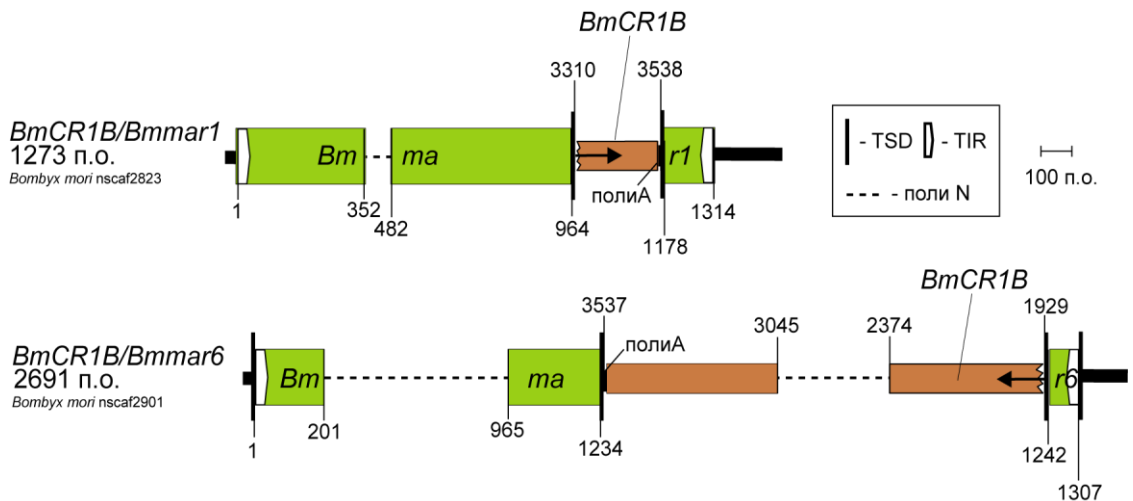


Рисунок 3.5. Структура химерных последовательностей *BmCR1B/MLF* из генома *B. mori*. Цифры внизу рисунка соответствуют позициям нуклеотидной последовательности *MLF*; цифры над рисунком соответствуют позициям нуклеотидной последовательности *BmCR1B*, стрелки указывают направление ориентации последовательности. Аббревиатура: п.о.- пары оснований, TIR - terminal inverted repeat (концевые инвертированные повторы), TSD - target site duplication (дупликация сайта – мишени).

Химерный элемент *BmCR1B/Bmmar1* содержит в своей структуре полноразмерный *Bmmar1* ДНК транспозон, со встроенным в него делетированным *BmCR1B* non-LTR ретротранспозоном. Встроенный фрагмент *BmCR1B* элемента имеет размер 237 п.о. и фланкирован TSD (ЦТЦТГА...ЦТЦТГА).

Химерный элемент *BmCR1B/Bmmar6* содержит полноразмерную копию *Bmmar6* *MLF* со встройкой *BmCR1B* элемента в противоположной относительно *Bmmar6* ориентации. Оба элемента в составе конструктора *BmCR1B/Bmmar6* содержат TSDs, что позволяют предположить относительно недавнюю активность *Bmmar6* *MLF* и *BmCR1B* non-LTR ретротранспозона.

3.2. *Mariner*-like TIR ДНК транспозоны в геномах представителей отряда *Lepidoptera*

На первом этапе исследования был подтвержден ГП *CR1B* non-LTR ретротранспозонов между представителями семейств *Lycaenidae* и *Bombycidae* и в геноме *B. mori* были выявлены две химерных конструкции, содержащих последовательности *CR1B* элемента, встроенные в *mariner*-like TIR ДНК транспозоны. Вышеописанные факты позволили предположить возможное участие *MLF* ДНК транспозонов в ГП *CR1B* элементов. Для проверки данной гипотезы был

проведен анализ распространения и разнообразия *MLE* TIR ДНК транспозонов в геномах представителей класса Insecta и отряда Lepidoptera.

К *MLEs* относятся ДНК транспозоны семейств *mariner*, *ludens* и *mori* суперсемейства *Tc1/mariner*. Представители *MLEs* обнаружены в геномах Растений, Животных, Простейших и Грибов (Plasterk *et al.*, 1999). Согласно современной классификации в составе семейства *mariner* представлено 15 подсемейств: *DTTMar* (*mariner*), *DTTMarAtl* (*atlantis*), *DTTMarBRI* (*briggsae*), *DTTMarCAP* (*capitata*), *DTTMarCec* (*cecropia*), *DTTMarCRI*, *DTTMarELE* (*elegans*), *DTTMarGGS*, *DDTMarIRR* (*irritans*), *DDTMarLin* (*lineata*), *DTTMarMau* (*mauritiana*), *DTTMarMel* (*mellifera*), *DTTMarROS* (*rosa*), *DTTMarUrt*, *DTTMarVer* (*vertumnana*). Внутри семейства *ludens* выделено два подсемейства, а внутри семейства *mori* одно подсемейство (*DTTMOR*) (Rouault *et al.*, 2009).

Наиболее изученным представителем отряда Lepidoptera является тутовый шелкопряд *B. mori*. Ранее в геноме *B. mori* было выявлено шесть *MLE* ДНК транспозонов, принадлежащих к двум семействам: *mariner* и *mori* (Robertson, Asplund, 1996; Robertson, Walden, 2003; Kumaresan, Mathavan, 2004). Элементы *Bmmar1* и *Bmmar6* относятся к семейству *mori*, элементы *Bmmar3* и *Bmmar4* к подсемейству *mellifera*, семейства *mariner*, а элементы *Bmmar4* и *Bmmar5* к подсемейству *cecropia*, семейства *mariner*. Анализ элементов *Bmmar1-6*, представленных в базе данных GenBank, показал, что данные элементы значительно дивергировали и сходство их аминокислотных последовательностей находится в диапазоне 11-36%. Ни одна из изученных копий *Tc1/mariner* элементов не содержит функционально активную транспозазу.

На основании данных представленных в литературе ожидается, что в геномах представителей отряда Lepidoptera будут выявлены представители как минимум двух семейств *MLE* ДНК транспозонов: *mariner* и *mori*.

3.2.1. Поиск *MLE* ДНК транспозонов в геномах представителей класса Insecta

Для изучения распространения и разнообразия *MLE* ДНК транспозонов в геномах представителей класса Insecta и отряда Lepidoptera с помощью программы BLAST (алгоритмы blastn и tblastp) был проведен поиск данных элементов в базе данных GenBank (Altschul *et al.*, 1990). В общей сложности, проведенный поиск

позволил выявить 85 последовательностей *MLE* элементов различной длины, 27 из которых представлены в геномах представителей отряда *Lepidoptera* (Приложение 2).

Полученные последовательности были использованы для дальнейшего сравнительного и филогенетического анализов. Множественное выравнивание полученных последовательностей осуществлялось при помощи программ Vector NTI 11.0 и Unipro UGENE. Выравнивания использовались для дальнейшего филогенетического анализа. В качестве алгоритма для построения филогенетического дерева на основе нуклеотидных последовательностей *MLE* элементов был применен метод соединения ближайших соседей (NJ) программы MEGA 5.0 (Tamura *et al.*, 2011) (Рисунок 3.6). На данном древе представлены последовательности *MLE* элементов из геномов представителей класса *Insecta* и отряда *Lepidoptera*, а также *MLE* элементы подсемейств *briggsae*, *atlantis*, *elegans*, *DTTmariner* и *DTTMarGGS* из геномов других групп живых организмов.

На филогенетическом древе выделяются 15 кластеров, соответствующих 15 группам *MLE* элементов. Наиболее многочисленными и разнообразными являются кластеры, соответствующие подсемействам *cecropia*, *mauritiana* и *mellifera*. Кластер *cecropia* составляют 19 элементов из 13 видов, принадлежащих восьми различным семействам отряда *Lepidoptera* и одному семейству отряда *Trichoptera*: *Papilionidae*, *Nymphalidae*, *Bombycidae*, *Saturniidae*, *Sphingidae*, *Arctiidae*, *Elachistidae*, *Pyrallidae* и *Hydroptilidae*. Внутри кластера выделяются две подгруппы (коэффициенты поддержки 93% и 94%), однако выделение данных групп не согласуется с филогенией видов хозяев, так обе группы содержат представителей семейства *Saturniidae* (Рисунок 3.6, Приложение 2).

Кластер *mauritiana* представлен 6 последовательностями из геномов насекомых шести различных отрядов: *Hymenoptera*, *Hemiptera*, *Coleoptera*, *Diptera* и *Lepidoptera*. В кластере *mellifera* выявлены элементы из геномов трех отрядов: *Trichoptera* (*Hydroptilidae*), *Diptera* (*Tabanidae*) и *Lepidoptera* (*Saturniidae*, *Yponomeutidae*, *Pyrallidae*, *Bombycidae*). Среди элементов кластеров *mauritiana* и *mellifera* невозможно выделить подгруппы.

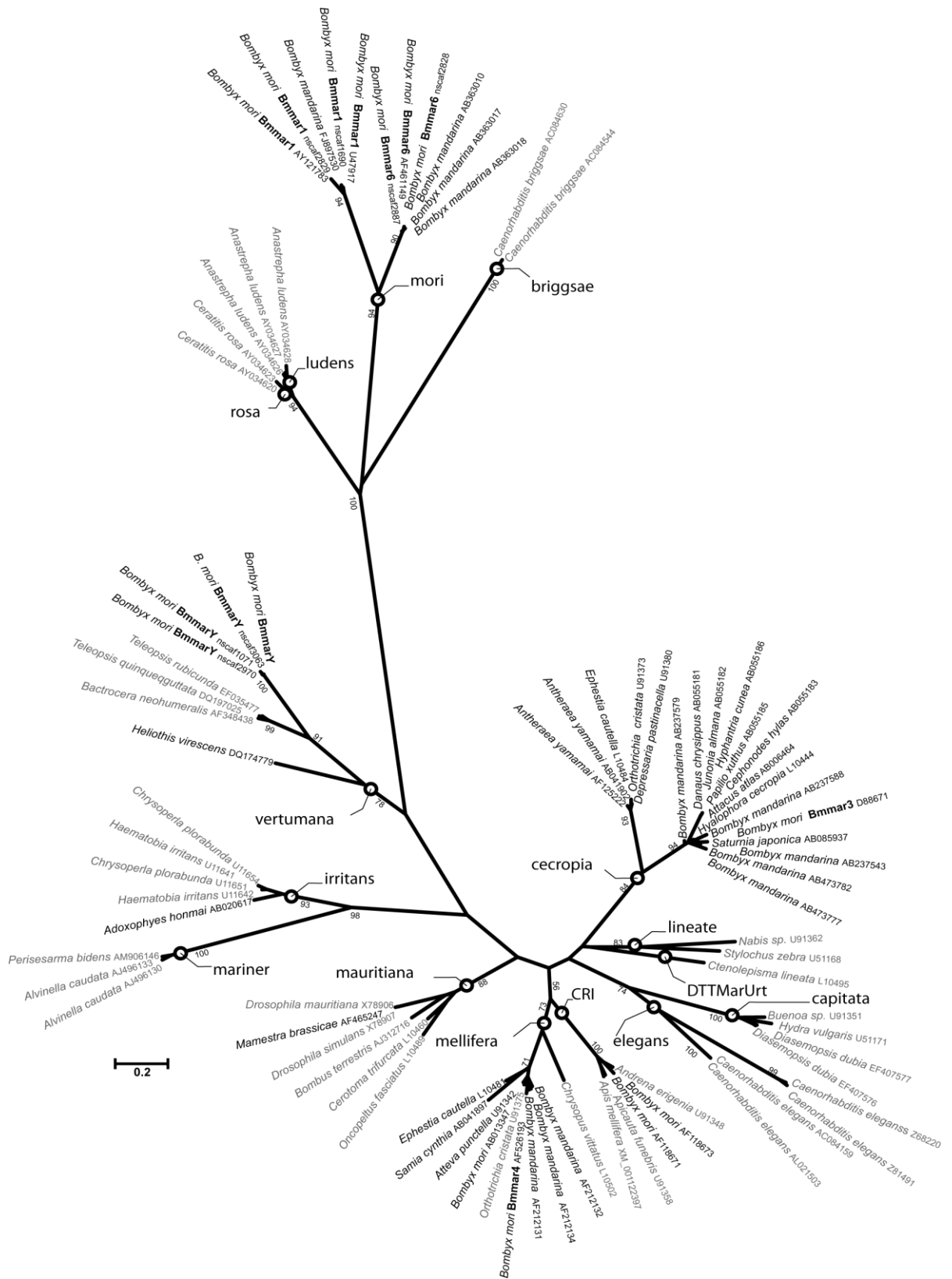


Рисунок 3.6. Бескорневое филогенетическое древо, реконструированное на основе частичных нуклеотидных последовательностей транспозазы (450 п.о.) *MLE* ДНК транспозонов методом объединения ближайших соседей (NJ) в программе MEGA 5.0. Для оценки достоверности топологии был использован бутстреп-тест (1000 репликаций). Значения коэффициентов поддержки бутстреп-теста менее 50% не показаны. Элементы, выявленные в геномах представителей отряда Lepidoptera, выделены жирным шрифтом. Номера нуклеотидных последовательностей элементов из базы данных GenBank приведены после названия вида.

В кластерах *vertumana* и *irritans* выявлено по одному элементу из геномов представителей отряда Lepidoptera: *Heliothis virescens* и *Adoxophyes honmai*, соответственно. Кроме отряда Lepidoptera элементы *vertumana* выявлены в геномах представителей отряда Diptera, а элементы *irritans* в геномах представителей отрядов Diptera и Neuroptera.

Распространение элементов кластера *mori*, по-видимому, ограничено исключительно геномами представителей рода *Bombyx* (Lepidoptera).

Кластеры *rosa*, *capitata*, *lineata*, *DTTMarUrt* представляют собой немногочисленные группы элементов и не были выявлены в геномах представителей отряда *Lepidoptera*. В геномах представителей класса Insecta к настоящему моменту не выявлено элементов подсемейств *briggsae*, *atlantis*, *elegans*, *DTTmariner* и *DTTMarGGS*. Группы *DTTmariner* и *DTTMarGGS* являются относительно новыми и получили статус подсемейств в результате автоматической классификации (Rouault *et al.*, 2009). Подсемейства *elegans* и *briggsae*, по-видимому, образованы исключительно элементами из геномов *C. elegans* и *C. briggsae* (Witherspoon, Robertson, 2003; Rouault *et al.*, 2009).

Таким образом, в геномах представителей класса Insecta представлено 13 подсемейств *mariner* и по одному подсемейству семейств *ludens* и *mori*. Разнообразие *MLEs* в геномах представителей отряда Lepidoptera ограничивается шестью подсемействами *mariner* (*DTTMarCRI*, *cecropia*, *mauritiana*, *mellifera*, *vertumana*, *irritans*) и семейством *mori*.

3.2.2. Поиск *MLE* ДНК транспозонов в геноме *B. mori*

Анализ генома тутового шелкопряда *B. mori* на наличие представителей *MLE* TIR ДНК транспозонов был проведен с помощью приложений Query Designer и НММ пакета программ UGENE (Eddy, 1998; Okonechnikov *et al.*, 2012). Для поиска элементов была сконструирована схема, включающая использование HMMER2 профиля. HMMER2 профиль был построен на основании выравнивания последовательностей *MLE* ДНК транспозонов из геномов насекомых. В результате проведенного анализа в геноме *B. mori* было выявлено семь *MLE* элементов, шесть из которых (*Bmmar1-6*) были ранее описаны в данном геноме (Robertson, Asplund, 1996; Robertson, Walden, 2003; Kumaresan, Mathavan, 2004). Данный поиск

позволил выявить новый элемент, обозначенный как *BmmarY*. Структура этого элемента была проанализирована с помощью программ BlastN, BlastP и ORF Finder и представлена на рисунке 3.7.

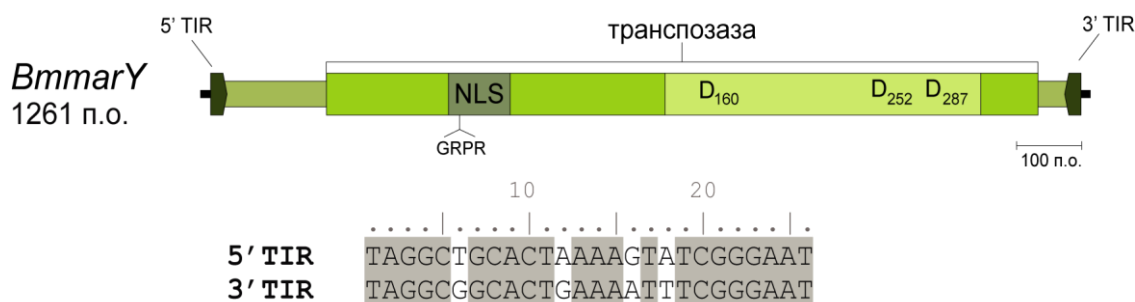


Рисунок 3.7. Структурная организация *BmmarY* ДНК транспозона из генома *B. mori* и нуклеотидные последовательности инвертированных повторов данного элемента. Аббревиатура: GRPR – аминокислотная последовательность глицин-аргинин-пролин-аргинин; NLS – nuclear location signal - сигнал ядерной локализации; TIR – terminal inverted repeat – концевые инвертированные повторы; п.о. – пары оснований.

Элемент *BmmarY* представляет собой типичный *mariner*-like ДНК транспозон. Данный элемент имеет длину 1,287 п.о. и содержит концевые инвертированные повторы длиной 26 п.о., сходство между которыми составляет 84.6%. Тело элемента содержит единственную открытую рамку считывания (1,050 п.о.), кодирующую фермент перемещения – транспозазу. Транспозаза *BmmarY* имеет характерный D,D(35)D мотив каталитического домена и два мотива сигнала ядерной локализации (GRPR и NLS). Поиск по базе данных EST *B. mori*, доступной на ресурсе SilkDB, проведенный с помощью программы BLASTN не выявил транскриптов транспозазы *BmmarY*.

Таким образом, можно предположить, что элемент *BmmarY* не экспрессируется в геноме *B. mori*. Распространенность *BmmarY* элемента, выраженная в числе полноразмерных копий на гаплоидный геном *B. mori*, составила 100 копий/гаплоидный геном. На основании консенсусной последовательности *BmmarY* были сконструированы праймеры для экспериментального поиска данного элемента в геномах других видов отряда Lepidoptera.

3.2.3. Экспериментальный поиск *MLE* ДНК транспозонов в геномах представителей отряда *Lepidoptera*

Для экспериментального поиска *MLE* ДНК транспозонов в геномах представителей *Lepidoptera* были выбраны 60 видов, 15-ти подсемейств из 8 суперсемейств (Таблица 3.1). Основываясь на результатах анализа разнообразия *MLE* ДНК транспозонов в геномах представителей *Insecta*, *Lepidoptera* и *B. mori*, было сделано предположение о том, что в результате экспериментального поиска могут быть выявлены элементы 7-ми подсемейств: *DTTMarCRI*, *cecropia*, *mauritanica*, *mellifera*, *vertumana*, *irritans* и *mori*. Поиск данных подсемейств проводился с помощью ПЦР амплификации с дегенеративной парой праймеров MAR-124F и MAR-276R, специфичной к последовательности транспозазы *MLEs* (Robertson, 1993) (Приложение 1).

Продукт ожидаемой длины (~ 450 п.о.) был получен для 36 из 60 проанализированных видов отряда *Lepidoptera* (Таблица 3.1). Для 25 видов, относящихся к семействам *Lycaenidae*, *Riodinidae* и *Bombycidae* продукты ПЦР амплификации были клонированы в плазмидном векторе. В целом было получено 190 клонов и для каждого из них была установлена нуклеотидная последовательность. 145 клонов из 18 видов содержали фрагменты, характеризующиеся высоким сходством с транспозазой *MLEs*. Данные элементы получили шестибуквенные названия: *Ocamar13* - *Oberthuria caeca mariner*-like элемент клон **13**.

Полученные последовательности транспозазы были выложены в базу данных GenBank под номерами: HQ606331–HQ606343, HQ606345–HQ606368, HQ606371–HQ606381, HQ606402, HQ606410, и HQ606415–HQ606462.

3.2.4. Анализ *MLE* ДНК транспозонов в геномах представителей отряда *Lepidoptera*

Сравнительный анализ последовательностей выявил высокую гетерогенность последовательностей транспозазы *MLEs*. Для дальнейшего филогенетического анализа были использованы последовательности фрагмента гена транспозазы, полученные в рамках экспериментального поиска в геномах представителей отряда *Lepidoptera* и обнаруженные в результате

биоинформатического поиска в геномах представителей класса Insecta. Кроме того, для более детального анализа были использованы последовательности элементов подсемейств *briggsae*, *atlantis*, *elegans*, *DTTmariner* и *DTTMarGGS*, ранее не обнаруженные в геномах представителей класса Insecta. Филогенетический анализ нуклеотидных последовательностей фрагмента гена транспозазы *MLE* элементов осуществлялся методом максимального правдоподобия (ML) в программе PhyML 3.0. Для оценки достоверности топологии использовался бутстреп-тест (1000 репликаций) (Felsenstein, 1985) (Рисунок 3.8).

На полученном филогенетическом древе представлено 17 кластеров *MLEs*: *DTTMarGGS*, *ludens*, *rosa*, *mori*, *vertumana*, *mauritiana*, *DTTMarCRI*, *mellifera*, *cecropia*, *DTTMarUrt*, *lineata*, *elegans*, *briggsae*, *capitata*, *irritans*, *atlantis* и *DTTmariner* (Рисунок 3.8). *MLE* элементы из геномов представителей отряда Lepidoptera полученные в результате экспериментального поиска обнаружены в 5-ти кластерах: *cecropia*, *mellifera*, *mauritiana*, *mori* и *vertumana*. Основная масса последовательностей выявленных в данной работе распределилась между кластерами *mauritiana*, *cecropia* и *mellifera* (Рисунок 3.8).

Кластер *cecropia*. Кластер *cecropia* является одним из наиболее многочисленных. Элементы данного кластера широко представлены в геномах 10 из 25 изученных видов отряда Lepidoptera: *Oberthuria caeca*, *Quentalia chromana*, *Erebia theano*, *Narathura japonica*, *Brangas neora*, *Scolitantides orion*, *Maculinea teleius*, *Maculinea nausithous*, *Euselasia chrysippe* и *Emesis lucinda* (Рисунок 3.8).

Среди элементов кластера *cecropia* можно выделить две группы (коэффициенты поддержки 100% и 99%) (Рисунок 3.8). В первую группу попали следующие элементы: *Vmmar5* (*B. mori*, Bombycidae) и элементы, выявленные исключительно в геномах представителей семейства Lycaenidae. Внутри данной группы не представляется возможным выделить более низкие классификационные единицы из-за очень высокой дивергенции последовательностей. Вторая группа состоит из двух подгрупп, в первую из которых попали исключительно копии элемента *Vmmar3* из генома *B. mori* (коэффициент поддержки 97%), а во вторую элементы из геномов представителей отрядов Riodinidae, Bombycidae и Lycaenidae (коэффициент поддержки 79%). Элементы трех данных семейства на филогенетическом древе образуют ветви согласно их видовой принадлежности.

Кластер *mellifera*. Кластер *mellifera* также является многочисленным и разнообразным кластером *MLEs* в геномах представителей отряда Lepidoptera. Элементы данного кластера представлены в геномах 10 видов, относящихся к 3 семействам: Bombycidae, Lycaenidae и Riodinidae (Рисунок 3.8). В данном кластере не представляется возможным выделить какие-либо подгруппы из-за очень высокой гомологии последовательностей.

Кластер *mauritiana*. В кластере *mauritiana* выявлены элементы исключительно из видов семейства Bombycidae: *Bombyx huttoni*, *Oberthuria caesa* и *Ernolatia moorei* (Рисунок 3.8). Кроме того, следует отметить, что данные элементы образуют отдельную ветвь внутри кластера и значительно отличаются по последовательностям от элементов семейства *mauritiana*, описанных ранее.

Кластер *vertumana*. В кластере *vertumana* выделяется несколько филогенетических групп. В первую группу попадают элементы подсемейства *vertumana* из геномов насекомых *Bactrocera neohumeralis*, *Teleopsis rubicunda* и *Teleopsis quinqueguttata* (отряд Diptera), описанные в литературе (коэффициент поддержки 100%), во вторую – элементы из геномов представителей родов *Bombyx* (Bombycidae) и *Maculinea* (Lycaenidae) (коэффициент поддержки 100%) (Рисунок 3.8). Элемент *BmmarY*, выявленный в рамках биоинформатического анализа генома *B. mori*, относится к подсемейству *vertumana*. Элементы из геномов представителей рода *Maculinea* имеют высокое сходство с *BmmarY* и получили следующие названия: *MtemarY.1* - *Maculinea teleius* элемент *BmmarY* клон 1). Парное сравнение нуклеотидных последовательностей кластера *vertumana* из геномов представителей родов *Bombyx* и *Maculinea* показало очень высокую гомогенность последовательностей как внутри видов, так и в межвидовых сравнениях.

Отдельную ветвь в данной группе составляют элементы из одного вида – *Pseudolucia collina* (Lycaenidae) (Рисунок 3.8). Последовательности данных элементов значительно отличаются от других элементов кластера *vertumana*. Возможно, данная группа представляет собой новое подсемейство *MLEs*.

В данную группу также попадает элемент из генома совки *Heliothis virescens* (Noctuidae), который изначально был отнесен к подсемейству *irritans* (Ren *et al.*, 2006). Данные, полученные в рамках филогенетического анализа, подтверждают,

что данный элемент относится к подсемейству *vertumana*, а не к подсемейству *irritans*.

Кластер *mori*. Изначально считалось, что элементы семейства *mori* (*Bmmar1* и *Bmmar6*) представлены исключительно в геномах представителей семейства *Bombycidae*. В рамках экспериментального этапа исследования элементы подсемейства *mori* были выявлены в геномах представителей рода *Maculinea* (семейство *Lycaenidae*). Сравнение нуклеотидных последовательностей элементов из геномов представителей рода *Maculinea* показало высокое сходство данных элементов с *Bmmar1* из генома *B. mori*. Данные элементы были названы следующим образом: *Mtemar1.3* - *Maculinea teleius* элемент *Bmmar1* клон 3. Элементы *Bmmar6* образуют отдельную ветвь в данном кластере, данные элементы не были обнаружены в геномах других видов отряда *Lepidoptera* (Рисунок 3.8).

3.2.5. Эволюция *MLE* ДНК транспозонов в геномах представителей отряда *Lepidoptera*

Согласно теории вертикальной эволюции должно наблюдаться соответствие между филогенией видов, основанной на последовательностях МГЭ, и филогенией, реконструированной на основе последовательностей молекулярных маркеров (функциональных генов *EF1 α* , *COI*, *COII* и других). Это означает, что мобильные элементы одного семейства должны образовывать ветви согласно видовой принадлежности. Элементы кластеров *cecropia*, *mellifera*, *mauritiana*, *irritans* и *DTTMarCRI*, по-видимому, эволюционировали в геномах представителей отряда *Lepidoptera* путем вертикального наследования (Рисунок 3.9).

Исключение составляют элементы кластеров *vertumana* и *mori*. Последовательности *BmmarY* элементов из геномов *Bombyx* и *Maculinea* попадают в одну ветвь кластера *vertumana* (Рисунок 3.9). Парное сравнение данных последовательностей показало высокую гомологию как внутри видов, так и в межвидовых сравнениях. Аналогично последовательности *Bmmar1* из геномов представителей родов *Bombyx* и *Maculinea* попадают в одну ветвь кластера *mori* (Рисунок 3.9). Как и в случае кластера *vertumana*, элементы кластера *mori* показали при попарных и множественных выравниваниях высокую гомогенность как внутривидовую, так и межвидовую.

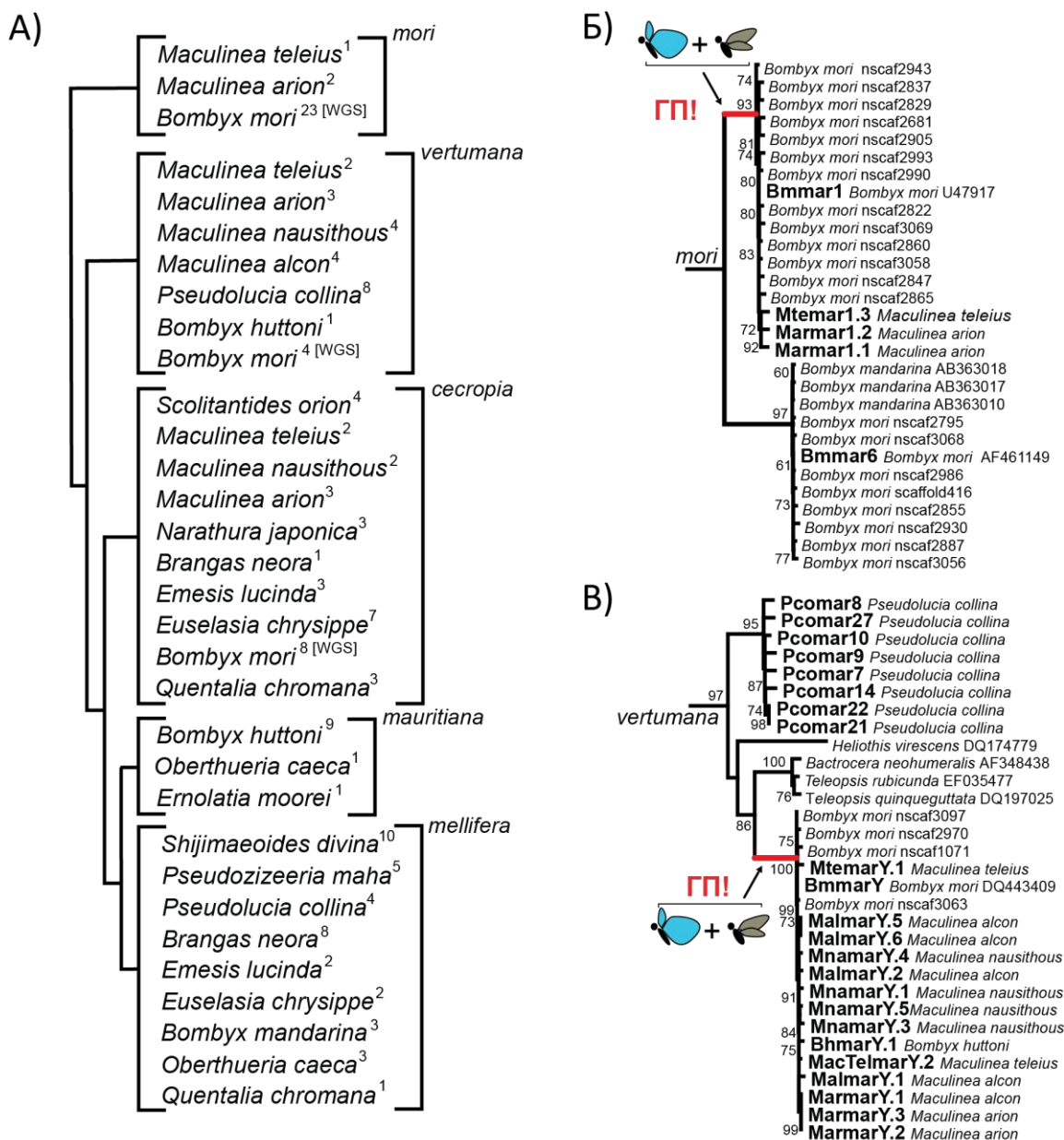


Рисунок 3.9. Схематичное представление филогенетического дерева, реконструированного на основе последовательностей транспозазы *mariner*-like элементов (*MLEs*). А) Иллюстрирует все подсемейства *MLEs* выявленные в геномах изученных видов отряда Lepidoptera. Число последовательностей, использованных в анализе, указано после названия вида. Б) и В) Детальные представления ветвей *mori* и *vertumana*. Бабочка голубого цвета обозначает бабочек голубянок семейства Lycaenidae, бабочка коричневого цвета – молей семейства Bombycidae. Аббревиатура: ГП – горизонтальный перенос; WGS - whole genome sequencing (полногеномное секвенирование).

Таким образом, элементы большинства подсемейств *MLEs* эволюционировали в геномах изученных видов отряда Lepidoptera путем вертикального наследования. Элементы *Bmmar1* (*mori*) и *BmmarY* (*vertumana*) из геномов представителей отряда Lepidoptera характеризуются следующими особенностями: 1) неравномерное распространение (элементы представлены в

геномах представителей родов *Bombyx* и *Maculinea*); 2) высокое сходство последовательностей между эволюционно удаленными группами; 3) несоответствие филогении данных элементов и филогении видов. Вышеописанные факты могут являться свидетельствами ГП элементов *Bmmar1* и *BmmarY* между представителями родов *Bombyx* и *Maculinea*.

Однако, сравнения последовательностей длиной в 450 п.о. недостаточно для достоверного разрешения эволюционных взаимоотношений элементов из геномов представителей родов *Bombyx* и *Maculinea*. Для подтверждения ГП *Bmmar1* и *BmmarY* необходимо проанализировать разнообразие и распространение полных последовательностей данных элементов в геномах представителей отряда *Lepidoptera*.

3.2.6. Экспериментальный поиск *Bmmar*-like элементов в геномах представителей отряда *Lepidoptera*

Для изучения разнообразия и эволюции полных *MLEs* был проведен поиск элементов *Bmmar* (*Bmmar*-like элементов) в геномах представителей родов *Maculinea*, *Bombyx* и некоторых других групп отряда *Lepidoptera*. Поиск проводился методом ПЦР амплификации с использованием пар праймеров, специфических к последовательностям TIRs элементов *Bmmar1*, *Bmmar3*, *Bmmar6* и *BmmarY*. Праймеры специфичные к TIRs элементов *Bmmar1*, *Bmmar3* и *Bmmar6* были взяты из литературных данных (Robertson, Asplund, 1996; Robertson, Walden, 2003; Kumaresan, Mathavan, 2004), праймеры специфичные к TIRs элемента *BmmarY* были сконструированы в рамках данного исследования (Приложение 1). В результате изучения генома *B. mori* не были выявлены полноразмерные копии элементов *Bmmar2*, *Bmmar4* и *Bmmar5*. Отсутствие полноразмерных последовательностей не позволило сконструировать специфические пары праймеров и провести ПЦР поиск данных элементов в геномах представителей отряда *Lepidoptera*.

Поиск элементов *Bmmar1*, *Bmmar3*, *Bmmar6* и *BmmarY* проводился в геномах 28 представителей отряда *Lepidoptera* (Таблица 3.2). Поскольку предполагаемый горизонтальный перенос произошел между родами *Maculinea* и *Bombyx*, для поиска *Bmmar* элементов были выбраны четыре вида рода *Maculinea* (*M.alcon*, *M.arion*,

M. nausithous и *M. teleius*) и три вида рода *Bombyx* (*B. mandarina*, *B. mori* и *B. huttoni*). Для оценки равномерности распространения *Bmmar* элементов в геномах представителей отряда Lepidoptera было выбрано несколько видов семейства *Lycaenidae*, к которому принадлежит род *Maculinea* (*Scolitantides orion*, *Shijimaoides divina*, *Pseudolucia collina*, *Pseudozizeeria maha*, *Narathura japonica* и *Brangas neora*), а также представители семейства *Bombycidae*, близких к роду *Bombyx* (*Oberthueria caeca*, *Quentalia chromana*). В качестве контроля были взяты представители семейств Riodinidae, Nymphalidae, Drepanidae, Callidulidae, Lymantriidae, Noctuidae, Sphingidae, Saturniidae (Таблица 3.2).

Таблица 3.2. Список исследованных видов отряда Lepidoptera, результаты ПЦР анализа элементов *Bmmar1*, *Bmmar3*, *BmmarY*, и *Bmmar6*.

Супер-семейство	Семейство	Вид	Результаты ПЦР амплификации			
			<i>Bmmar3</i>	<i>Bmmar1</i>	<i>BmmarY</i>	<i>Bmmar6</i>
Papilionoidea	Lycaenidae	<i>Scolitantides orion</i>	+	-	-	-
		<i>Shijimaoides divina</i>	+	-	-	-
		<i>Pseudozizeeria maha</i>	+	-	-	-
		<i>Maculinea teleius</i>	+	+	+	+
		<i>Maculinea nausithous</i>	+	+	+	+
		<i>Maculinea arion</i>	+	+	+	+
		<i>Maculineaalcon</i>	+	+	+	+
		<i>Pseudolucia collina</i>	+	-	-	-
		<i>Narathura japonica</i>	+	-	-	-
		<i>Brangas neora</i>	+	-	-	-
	Riodinidae	<i>Emesis lucinda</i>	+	-	-	-
		<i>Euselasia chrysippe</i>	+	-	-	-
		<i>Theope virgilius</i>	+	-	-	-
		<i>Caria rhacotis</i>	+	-	-	-
	Nymphalidae	<i>Araschnia levana</i>	+	-	-	-
		<i>Melitaea phoebe</i>	+	-	-	-
		<i>Erebia theano</i>	+	-	-	-

Таблица 3.2 (продолжение). Список исследованных видов отряда Lepidoptera, результаты ПЦР анализа элементов *Bmmar1*, *Bmmar3*, *BmmarY*, и *Bmmar6*.

Супер-семейство	Семейство	Вид	Результаты ПЦР амплификации			
			<i>Bmmar3</i>	<i>Bmmar1</i>	<i>BmmarY</i>	<i>Bmmar6</i>
Drepanoidea	Drepanidae	<i>Drepana sp.</i>	+	-	-	-
Calliduloidea	Callidulidae	<i>Callidula sp.</i>	+	-	-	-
Noctuoidea	Lymantriidae	<i>Lymantria dispar</i>	+	-	-	-
	Noctuidae	<i>Polia nebulosa</i>	+	-	-	-
Bombycoidea	Sphingidae	<i>Agrius cingulata</i>	+	-	-	-
	Saturniidae	<i>Oxytenis modestia</i>	+	-	-	-
	Bombycidae	<i>Bombyx mandarina</i>	+	+	-	+
		<i>Bombyx huttoni</i>	+	-	+	-
		<i>Bombyx mori</i>	+	+	+	+
		<i>Oberthueria caeca</i>	+	-	-	-
		<i>Quentalia chromana</i>	+	-	-	-

Положительный результат ПЦР реакции с праймерами специфичными к TIRs элемента *Bmmar3* были выявлены для всех 28 исследованных видов. Продукты, полученные в результате ПЦР амплификации, были клонированы и установлены их нуклеотидные последовательности. В общей сложности было получено 7 клонов *Bmmar3*-like *MLEs*. Анализ полученных последовательностей при помощи BLAST (алгоритмы blastn и tblastp) показал, что все клоны содержат последовательность транспозазы и имеют сходство с последовательностью *Bmmar3*. *Bmmar3*-like *MLEs* получили шестибуквенные названия: *Marmar3.2* - ***Maculinea arion*** ***Bmmar3***-like элемент клон **2**. Полученные последовательности *Bmmar3*-like элементов были выложены в базу данных GenBank под номерами: HQ606394-HQ606398, HQ606405 и HQ606412.

Полученные данные согласуются с результатами ПЦР амплификации с дегенеративными праймерами и подтверждают равномерное распространение элементов подсемейства *secropia* (*Bmmar3*-like) в геномах изученных видов отряда

Lepidoptera. Элементы *Bmmar3-like* являются примером *MLEs*, эволюция которых в геномах представителей отряда Lepidoptera происходила исключительно путем вертикального наследования.

Положительные сигналы ПЦР амплификации с праймерами специфичными к TIRs элементов *Bmmar1* и *Bmmar6* были получены для представителей семейства Bombycidae (*B. mori* и *B. mandarina*) и семейства Lycaenidae (*M. teleius*, *M. nausithous*, *M. arion* и *M. alcon*). В случае ПЦР реакции с праймерами специфичными к TIRs элемента *BmmarY* положительный результат был получен для представителей Bombycidae (*B. mori* и *B. huttoni*) и Lycaenidae (*M. teleius*, *M. nausithous*, *M. arion* и *M. alcon*). Негативные сигналы ПЦР амплификации с праймерами *Bmmar1-ITR*, *Bmmar6-ITR* и *BmmarY-ITR* были получены для оставшихся изученных видов. Таким образом, распространение элементов *Bmmar1-like*, *Bmmar6-like* и *BmmarY-like* среди представителей Lepidoptera ограничено родами *Bombyx* (Bombycidae) и *Maculinea* (Lycaenidae).

Полученные ПЦР продукты были клонированы и для 3 клонов *Bmmar1-like*, 2 клонов *Bmmar6-like* и 15 клонов *BmmarY-like* элементов были установлены нуклеотидные последовательности. Сравнение полученных последовательностей с известными мобильными элементами из GenBank при помощи различных алгоритмов BLAST показало, что все полученные клоны *Bmmar6-like* на самом деле представляют собой последовательности *Bmmar1-like* элементов. Данный факт, по-видимому, объясняется отсутствием *Bmmar6-like* элементов в геномах представителей семейства Lycaenidae, а положительный сигнал ПЦР связан с высокой дегенеративностью праймеров *Bmmar6-ITR*. Последовательности *Bmmar1-like* и *BmmarY-like MLEs* (или *Macmar1* и *MacmarY*) были выложены в базу данных GenBank под номерами: HQ606330, HQ606344, HQ606387 - HQ6063930, HQ606392, HQ606393, HQ606399-HQ606401, HQ606403, HQ606404, HQ606406-HQ606409, HQ606411, HQ606413 и HQ606414.

3.2.7. Неравномерное распространение *Bmmar-like* элементов в геномах представителей отряда Lepidoptera

Результаты ПЦР амплификации свидетельствуют о неравномерном распространении *Bmmar1-like* и *BmmarY-like* элементов в геномах изученных видов

отряда Lepidoptera. Поскольку ПЦР реакция является чувствительной и подвержена контаминации, для подтверждения результатов ПЦР была проведена дот-блот гибридизация. Анализ проводился на геномной ДНК 25 видов отряда Lepidoptera, в том числе *M. teleius*, *M. arion*, *B. mori* и *B. mandarina* (Рисунок 3.10).

Достоверные сигналы гибридизации с пробами специфичными к *Bmmar1* и *BmmarY* элементам были выявлены только для образцов ДНК представителей родов *Bombyx* и *Maculinea* (Рисунок 3.10). Слабые сигналы, выявленные для образцов *M. phoebe* и *E. theano*, в случае *Bmmar1* и *M. phoebe* и *E. theano*, *O. sculda*, *P. stubbendortii*, *C. glaucescens* в случае *BmmarY*, по-видимому, являются результатом перекрестной гибридизации проб с другими *MLE* элементами (Рисунок 3.10).

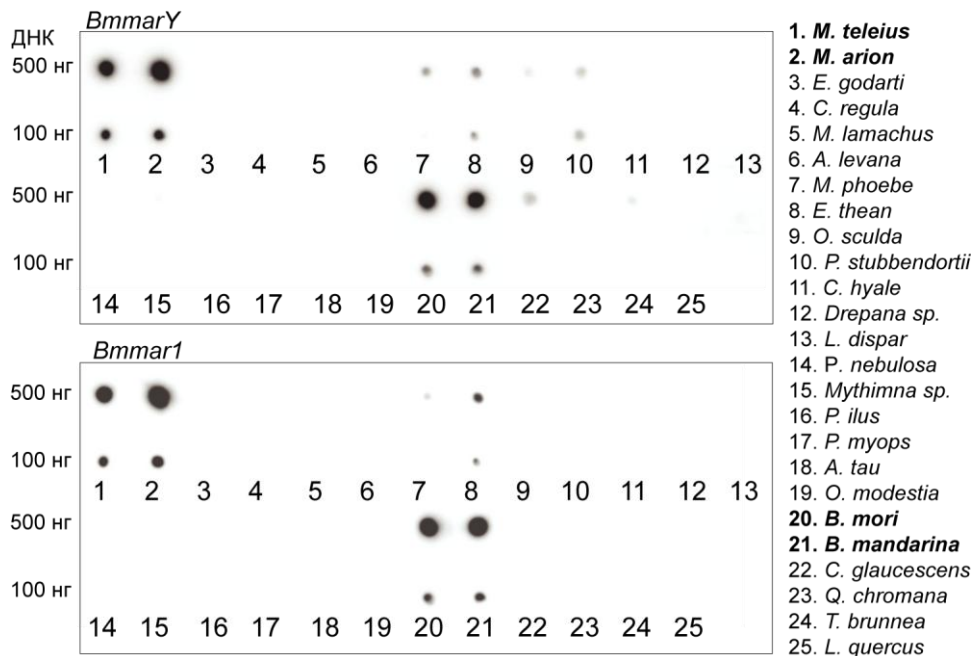


Рисунок 3.10. Результат дот-блот гибридизации образцов ДНК 25 видов бабочек и молей с пробами специфичными к элементам *Bmmar1*-like и *BmmarY*-like. Список использованных видов приведен справа, номера соответствуют номерам в таблице 3.1. В качестве зонда был использован P^{32} -меченный фрагмент последовательности обратной транскриптазы *CRIB* элемента из *M. teleius*.

Таким образом, результаты дот-блот гибридизации подтверждают результаты ПЦР амплификации и неравномерное распространение *Bmmar1*-like и *BmmarY*-like элементов в геномах изученных видов отряда Lepidoptera.

3.2.8. Анализ *Bmmar*-like элементов в геномах представителей отряда *Lepidoptera*

Сравнительный анализ последовательностей *Bmmar*-like выявил высокий уровень гомологии на межвидовом и внутривидовом уровнях для элементов *Bmmar1*-like и *BmmarY*-like (Таблица 3.3).

Уровень сходства последовательностей *BmmarY*-like элементов из геномов *Bombyx* и *Maculinea* является высоким (96.6%) и соответствует уровню сходства последовательностей внутри данных родов. Аналогичные результаты получены для *Bmmar1*-like элементов: сходство последовательностей из геномов представителей родов *Bombyx* и *Maculinea* составляет 92.3%. В случае *Bmmar3*-like элементов сходство последовательностей из геномов представителей *Bombyx* и *Maculinea* составляет 84.1%, а внутри родов уровень сходства составляет 90.1% и 93.9%, соответственно (Таблица 3.3).

Таблица 3.3. Сравнение нуклеотидных последовательностей различных *Bmmar*-like ДНК транспозонов.

Ген	Длина (п.о.)	Сходство (%)
<i>Bmmar1: Bombyx / Bombyx</i>	1251	96.0
<i>Bmmar1</i> -like: <i>Maculinea / Maculinea</i>	1227	93.4
<i>Bmmar1</i>-like: <i>Bombyx / Maculinea</i>	1130	92.3 (ГП)
<i>Bmmar3: Bombyx / Bombyx</i>	793	90.1
<i>Bmmar3</i> -like: <i>Maculinea / Maculinea</i>	820	93.9
<i>Bmmar3</i>-like: <i>Bombyx / Maculinea</i>	820	84.1 (вертикальная эволюция)
<i>BmmarY: Bombyx / Bombyx</i>	1248	97.9
<i>BmmarY</i> -like: <i>Maculinea / Maculinea</i>	1088	97.2
<i>BmmarY</i>-like: <i>Bombyx / Maculinea</i>	1085	96.6 (ГП)

Нуклеотидные последовательности *Bmmar*-like элементов были транслированы для анализа аминокислотных последовательностей транспозазы. На рисунке 3.11 представлено выравнивание последовательностей каталитических доменов транспозазы (DDE/D мотивов).

Все изученные последовательности содержат DDD мотив – один из двух вариантов мотивов, характерных для *mariner*-like элементов (Munoz-Lopez, Garcia-Perez, 2010). Между *Bmmar*-like элементами варьирует значение интервала между второй и третьей аминокислотами данного мотива (Рисунок 3.11).

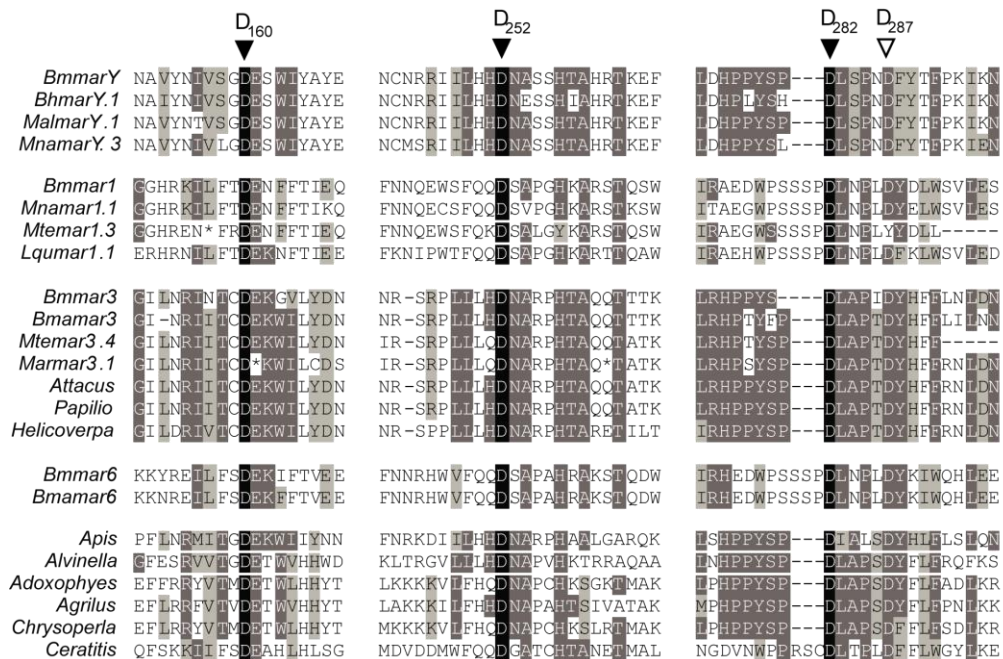


Рисунок 3.11. Выравнивание консенсусных аминокислотных последовательностей каталитического домена транспозазы различных групп *MLEs* из геномов *Lepidoptera*. Черным цветом выделены аминокислоты образующие DDD мотив.

Все транслированные последовательности содержали сдвиги рамки считывания в результате инсерции или делеции и единичные стоп-кодона. Таким образом, ни один из проанализированных элементов не содержал интактную последовательность гена транспозазы.

Для филогенетического анализа были использованы *Bmmar*-like элементы, выявленные экспериментальным путем в геномах представителей отряда *Lepidoptera* и биоинформатическим методом в геноме *B. mori* и генома представителей класса *Insecta*. Филогенетическое дерево на основе полноразмерных

последовательностей элементов было сконструировано с помощью алгоритма ML в программе PhyML 3.0. (Рисунок 3.12).

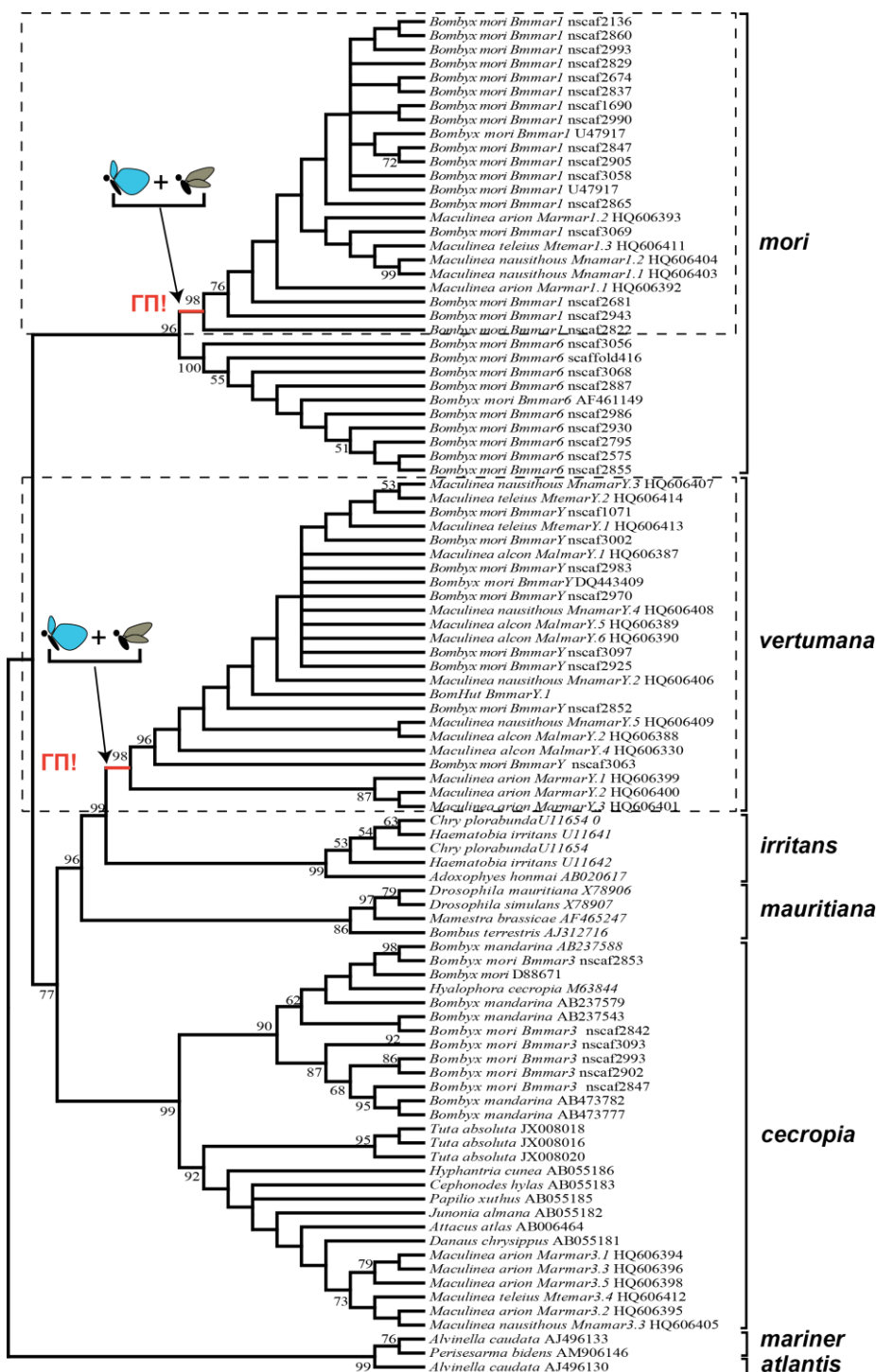


Рисунок 3.12. Филогенетическое древо, реконструированное на основе последовательностей полных *MLEs*, с помощью метода максимального правдоподобия (ML - Maximum Likelihood) в программе PhyML 3.0. В качестве внешней группы были использованы последовательности *MLEs* подсемейств *atlantis* и *mariner* из геномов представителей подкласса кольчатых червей и подкласса ракообразных. Для статистической оценки достоверности был использован бутстреп-тест (100 репликаций). Значения коэффициентов поддержки бутстреп-теста менее 50% не показаны. Бабочка голубого цвета обозначает бабочек голубянок семейства Lycaenidae, бабочка коричневого цвета – молей семейства Bombycidae.

В качестве внешней группы были использованы последовательности *MLE* элементов подсемейств *atlantis* и *mariner* из геномов представителей подкласса кольчатых червей и подкласса ракообразных. Для статистической оценки достоверности был использован бутстреп-тест (100 репликаций) (Рисунок 3.12).

На филогенетическом древе представлено пять кластеров: *mori*, *vertumana*, *mauritianana*, *irritans* и *cecropia*, характеризующихся высокой статистической поддержкой (Рисунок 3.12). Аналогично филогении, построенной на частичных последовательностях транспозазы, в случае полноразмерных элементов в кластере *cecropia* выделяется две группы. В первую группу (коэффициент поддержки 90%) распределились элементы из генома моли *Hyalophora cecropia* (суперсемейство *Bombycoidea*) и *Bmmar3*-like элементы из геномов представителей рода *Bombyx*. Во вторую группу (коэффициент поддержки 92%) распределились элементы из геномов 11 видов 7 семейств отряда *Lepidoptera* (*Gelechiidae*, *Arctiidae*, *Sphingidae*, *Papilionidae*, *Nymphalidae*, *Saturniidae* и *Lycaenidae*), в том числе и *Bmmar3*-like элементы из геномов *Maculinea*. Таким образом, последовательности *Bmmar3*-like элементов в кластере *cecropia* образуют ветви согласно видовой принадлежности.

В кластерах *mori* и *vertumana* представлены исключительно элементы из геномов представителей родов *Bombyx* и *Maculinea*. В кластере *mori* выделяются 2 группы, первая образована последовательностями *Bmmar6*-like исключительно из генома *B. mori*, а вторая *Bmmar1*-like элементами из геномов *Bombyx* и *Maculinea*.

Кластер *vertumana* сформирован последовательностями *BmmarY*-like элементов представителей родов *Bombyx* и *Maculinea*. Как в случае *BmmarY*-like, так и в случае *Bmmar1*-like элементов невозможно провести выделение отдельных ветвей соответствующих различным родам и видам. Таким образом, *Bmmar1*-like и *BmmarY*-like не образуют ветвей на филогенетическом древе, согласующихся с их видовой принадлежностью.

3.2.9. Подтверждение горизонтального переноса *Bmmar*-like элементов в геномах представителей отряда *Lepidoptera*

В отличие от non-LTR ретротранспозонов, для которых регистрируются лишь единичные случаи ГП, ГП ДНК транспозонов частое явление. Одним из

классических примеров ГП ДНК транспозонов является перенос элемента *mariner* между представителями класса Insecta (Maruyama, Hartl, 1991). Последовательности элементов *mariner* из геномов эволюционно удаленных видов *Zaprionus tuberculatus* и *D. mauritiana* характеризовались высоким сходством, превосходившим сходство последовательностей из близкородственных видов (*D. tsacasi* и *D. mauritiana*). Сравнение филогении на основе последовательностей *mariner* и филогении видов, построенной на основании консервативных последовательностей гена алкогольдегидрогеназы (АДГ), выявило несоответствие в позиции *Z. tuberculatus*. Таким образом, на основе высокого сходства последовательностей и несоответствия филогений был подтвержден ГП элемента *mariner* между *Z. tuberculatus* и *D. mauritiana* (Maruyama, Hartl, 1991). В дальнейших исследованиях были выявлены другие случаи ГП *MLEs* между геномами насекомых (Lohe *et al.*, 1995; Lampe *et al.*, 2003). В случае *MLE* элементов ГП является одним из ключевых этапов жизненного цикла, позволяющим им избежать полного исчезновения и внедриться в новый геном (Hartl *et al.*, 1997).

Высокое сходство последовательностей *BmmarY-like* и *BmmarI-like* элементов из геномов репродуктивно изолированных видов (*Bombyx* и *Maculinea*), их неравномерное распределение среди изученных видов отряда Lepidoptera и несоответствие филогении основанной на данных последовательностях филогении видов Lepidoptera являются свидетельствами ГП данных элементов.

Несмотря на то, что случай эволюции *BmmarY-like* и *BmmarI-like* удовлетворяет трем основным критериям ГП, описанным Schaack *et al.* (2010), современный уровень исследований требует использования параметрических методов анализа для подтверждения данного явления. Одним из таких методов является сравнение числа синонимичных замен на синонимичный сайт (dS). В случае, если МГЭ попал в геном в результате вертикального наследования величина dS в последовательности МГЭ должна соответствовать dS консервативных генов геномов хозяев. Если МГЭ был перенесен в геном горизонтально, то значение dS его последовательности МГЭ должно быть значительно ниже, чем dS консервативных генов (Silva, Kidwell, 2000; Ludwig *et al.*, 2008; Bartolome *et al.*, 2009; Mota *et al.*, 2010).

Для сравнительного анализа dS было использовано три группы последовательностей:

- 1) последовательности ядерных генов, наследование которых происходило вертикально. Последовательности генов *wg*, *EF1 α* и *H3* *B. mori* и *Maculinea* (*M. teleius*, *M. nausithous*, *M. arion* и *M.alcon*) были взяты из генетической базы данных GeneBank.
- 2) последовательности *BmmarY-like* и *Bmmar1-like* элементов из геномов *B. mori* и *Maculinea* (*M. teleius*, *M. nausithous*, *M. arion* и *M.alcon*), для которых предполагается ГП.
- 3) последовательности *Bmmar3-like* элементов из геномов *B. mori* и *Maculinea* (*M. teleius*, *M. nausithous*, *M. arion* и *M.alcon*), эволюция которых происходила вертикально.

Значения dS для последовательностей *wg*, *EF1 α* , *H3*, *BmmarY-like*, *Bmmar3-like*, *Bmmar1-like* для каждой из четырех пар видов (*B. mori* vs. *M. arion*, *B. mori* vs. *M. nausithous*, *B. mori* vs. *M. teleius*, *B. mori* vs. *M.alcon*) были подсчитаны в программе MEGA 5.0. Достоверность различия подсчитанных значений dS для транспозонов и ядерных генов была проанализирована с помощью точного теста Фишера (Рисунок 3.13).

Для пары *B. mori* и *M. arion* значение dS *Bmmar3-like* оказалось низким, разница между ним и значением dS ядерных генов оказалась недостоверной (Рисунок 3.13). Отсутствие достоверной разницы значений dS подтверждает вертикальное наследование *Bmmar3-like* элементов у *B. mori* и *M. arion*. В случае двух других пар (*B. mori* и *M. nausithous*, *B. mori* и *M. teleius*) разница значений dS *Bmmar3-like* элементов и ядерных генов оказалась достоверной ($P < 0.05$). Полученный результат ставит под сомнение единственность вертикального наследования в эволюции *Bmmar3-like* элементов *B. mori*, *M. nausithous* и *M. teleius*.

dS в случае *BmmarY-like* и *Bmmar1-like* элементов для всех исследованных пар была достоверно ниже значений dS ядерных генов (Рисунок 3.13).

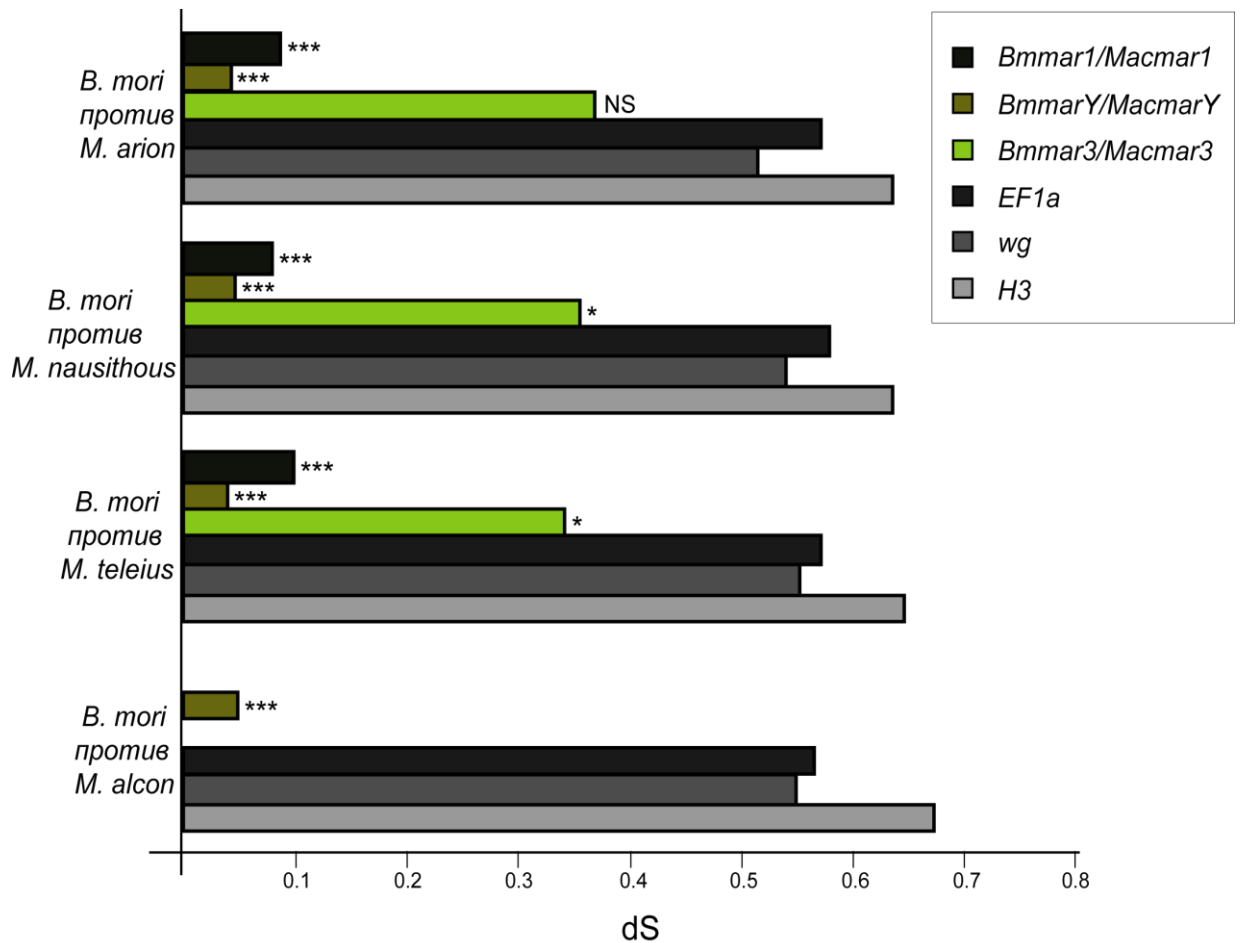


Рисунок 3.13. Значения dS для *mariner*-like элементов *Bmmar1*, *BmmarY* и *Bmmar3*, и ядерных генов фактора элонгации альфа (*EF1α*), *wingless* (*wg*) и гистона *H3* (*H3*) в попарных сравнениях между *B. mori* и видами *M. arion*, *M. nausithous*, *M. teleius*, и *M. alcon*. Результаты точного теста Фишера: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; NS – не достоверно (non-significant).

Полученные результаты подтверждают гипотезу о ГП *BmmarY*-like и *Bmmar1*-like элементов между бабочками Lycaenidae (род *Maculinea*) и молями Bombycidae (род *Bombyx*).

3.3. Совместный горизонтальный перенос *CR1B* non-LTR ретротранспозонов и *mariner*-like ДНК транспозонов

Процессы ГП *CR1B* non-LTR ретротранспозонов и *BmmarY*-like и *Bmmar1*-like *MLE* TIR ДНК транспозонов могли произойти независимо. Однако ГП элементов из двух различных групп МГЭ между определенными группами организмов и наличие химерных конструкций *CR1B/MLE* в геноме *B. mori* свидетельствуют о совместном переносе данных элементов. В случае если *MLEs*

принимали участие в ГП *CRIB* non-LTR ретротранспозонов между представителями отряда *Lepidoptera*, химерные конструкции аналогичные конструкциям, выявленным в геноме *B. mori*, должны присутствовать в геномах *Maculinea*. С целью подтверждения гипотезы об участии *MLE* ДНК транспозонов в ГП *CRIB* non-LTR ретротранспозонов был проведен экспериментальный поиск химерных последовательностей *CRIB/MLE* в геномах представителей отряда *Lepidoptera*.

3.3.1. Поиск химерных *CRIB/MLE* элементов в геномах представителей отряда *Lepidoptera*

Экспериментальный поиск химерных конструкций, содержащих последовательности *Bmmar*-like и *CRIB* элементов, в геномах представителей отряда *Lepidoptera* проводился с помощью ПЦР амплификации. Для реакции были сконструированы специфические пары праймеров, в каждой паре один из праймеров был комплементарен последовательности *Bmmar*-like элемента (*BmmarY*-like, *Bmmar1*-like или *Bmmar6*-like), а второй последовательности *CRIB* элемента (Приложение 1). Положительные сигналы ПЦР реакции были получены с единственной парой праймеров для видов *M. teleius*, *M. arion*, *B. mori* и *B. mandarina*.

Полученные фрагменты были использованы для приготовления минибиблиотеки клонов. Далее был проведен скрининг библиотек методом «colony lift assay» с целью выявления фрагментов, содержащих последовательности как *MLE*, так и *CRIB* элементов. Основная масса клонов содержала только фрагмент *MLE*, однако несколько клонов показали положительный сигнал гибридизации с пробамми на *CRIB*. По результатам секвенирования было выявлено три клонa, содержащих фрагменты последовательностей *CRIB*-like и *BmmarY*-like элементов: 2 клонa из генома *M. teleius* (*MacCRIB/MtemarY.1* и *MacCRIB/MtemarY.2*) и один клон из генома *M. arion* (*MacCRIB/MarmarY*) (Рисунок 3.14). Полученные последовательности *CRIB/MLE* из геномов *Maculinea* выложены в базу данных GenBank под номерами: JQ580972–JQ580974.

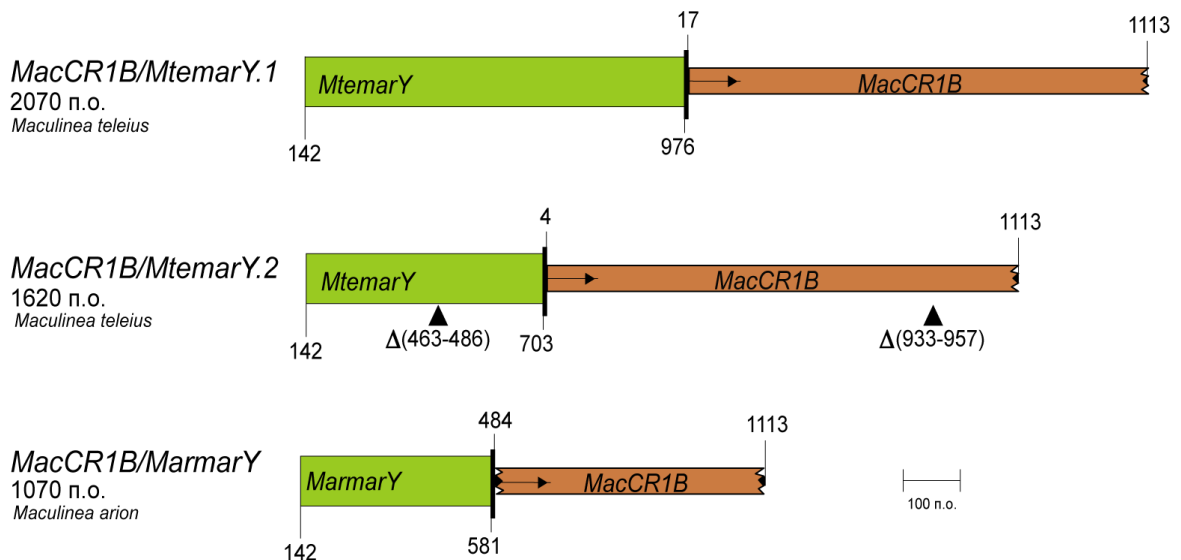


Рисунок 3.14. Структурная организация химерных элементов *CRIB/MLE* из геномов представителей рода *Maculinea*. Цифры внизу рисунка соответствуют позициям консенсусной последовательности *MLEs*; цифры над рисунком соответствуют позициям консенсусной последовательности *MacCR1B*. Черные треугольники обозначают делетированные участки, стрелки указывают направление ориентации последовательности.

Сравнительный анализ полученных последовательностей показал, что они значительно отличаются между собой и отличаются от химерных последовательностей, выявленных ранее в геноме *B. mori*. Анализ транслированных последовательностей полученных с помощью программы ORFFinder выявили один элемент *MacCR1B/MtemarY.1* (JQ580972) в случае которого встройка *CRIB* элемента не нарушила рамку считывания ДНК транспозона. Данная структура, возможно, представляют собой химерную конструкцию, участвовавшую в ГП (Рисунок 3.14).

3.3.2. *MLE* ДНК транспозоны в качестве векторов горизонтального переноса *CRIB non-LTR* ретротранспозонов

В результате проведенной работы был подтвержден факт ГП *CRIB non-LTR* ретротранспозонов между геномами представителей семейства Lycaenidae (род *Maculinea*) и семейства Bombycidae (родов *Bombyx* и *Oberthueria*) (Рисунок 3.15А).

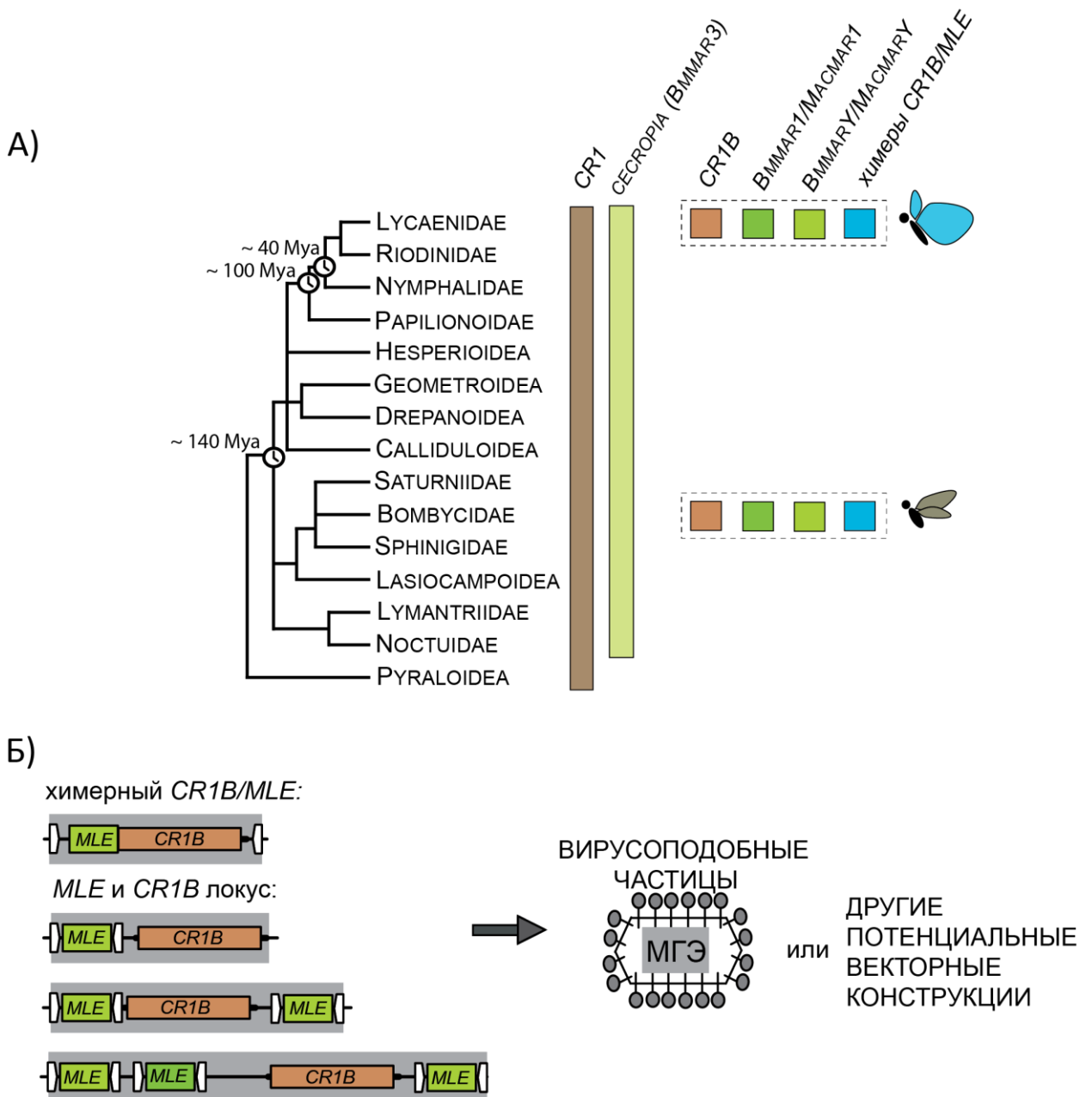


Рисунок 3.15. (А) Распространение *CR1* non-LTR ретротранспозонов, *MLE* ДНК транспозонов и химерных элементов *CR1B/MLE* в геномах представителей отряда Lepidoptera. Филогения видов отряда представлена согласно Tree of Life Web Project (<http://tolweb.org/>). (Б) Предполагаемая структура конструкций, участвовавших в горизонтальном переносе и механизм горизонтального переноса. Аббревиатура: МГЭ – мобильные генетические элементы, *MLEs* - *mariner* – like elements (*mariner*-like подобные элементы). Бабочка голубого цвета обозначает бабочек голубянок семейства Lycaenidae, бабочка коричневого цвета – молей семейства Bombycidae.

Установлен факт ГП *Bmmar1*-like и *BmmarY*-like ДНК транспозонов между геномами представителей семейства Lycaenidae (род *Maculinea*) и семейства Bombycidae (родов *Bombyx*) (Рисунок 3.15А). В геномах *B. mori* и *Maculinea* обнаружены химерные конструкции содержащие последовательности *CR1B* элементов, встроенные в последовательность *MLE* элементов (Рисунок 3.15А).

Вышеперечисленные результаты косвенно подтверждают участие ДНК транспозонов в ГП non-LTR ретротранспозонов.

В пользу данного сценария свидетельствует ряд фактов, известных из литературы:

1) *MLE* элементы обладают высокой способностью к ГП. Поскольку для перемещения *MLEs* ключевое значение имеет последовательность TIRs, встройки non-LTR ретротранспозонов не нарушающие TIRs вероятнее всего не повлияют на активность элемента. При условии нарушения последовательности транспозазы в результате встройки non-LTR ретротранспозона перемещение химерного конструкта может осуществляться за счет фермента транспозазы, наработанного с другой копии элемента (Feschotte, Pritham, 2007; Munoz-Lopez, Garcia-Perez, 2010) (Рисунок 3.15Б).

2) *MLE* элементы могут перенести non-LTR ретротранспозон из одного сайта в другой, если non-LTR ретротранспозон встроен в участок ДНК между двумя *MLE* элементами. Данный процесс осуществляется за счет того, что фермент транспозаза распознает два TIR, принадлежащих разным элементам, и вырезать всю последовательность ДНК, расположенную между ними (Рисунок 3.15Б). Для ГП между видами любой из вариантов химерных конструктов представленных на рисунке 3.15Б должен встроиться в промежуточного носителя – вектор (внутриклеточного паразита, вируса и тп.).

Полученные данные, тем не менее, не позволяют полностью исключить возможность того, что ГП *CRIB* non-LTR ретротранспозонов между геномами *Lycaenidae* и *Bombycidae* произошел независимо от перемещения *MLE* элементов. Альтернативным вариантом является ГП в виде РНК интермедиата. Данный вариант предполагает непосредственный захват и упаковку РНК интермедиата non-LTR ретротранспозона в вирусную частицу, какого либо вируса насекомых. При попадании в клетку вида реципиента происходит обратная транскрипция и non-LTR ретротранспозон может встроиться в новый геном. Таким образом, может осуществляться ГП между видами, подверженными заражению общими вирусами. Для некоторых вирусов была показана возможность упаковки различных РНК клетки-хозяина, в том числе РНК МГЭ, в собственные вирусные частицы (Routh *et*

al., 2012). Данный сценарий позволяет предположить и совместный ГП ДНК *MLE* элемента и РНК *CR1B* non-LTR ретротранспозона в упакованном виде в составе вирусоподобной частицы (Рисунок 3.15Б). Однако в настоящее время не получено результатов, доказывающих данный сценарий ГП МГЭ между геномами представителей семейств *Lycaenidae* и *Bombycidae*.

3.3.3. Механизм горизонтального переноса *CR1B* non-LTR ретротранспозонов и *MLE* ДНК транспозонов

Механизм ГП *CR1B* non-LTR ретротранспозонов и *MLE* ДНК транспозонов неизвестен. Для успешного ГП ДНК транспозонов и non-LTR ретротранспозонов требуется наличие вектора и возможности взаимодействия генома донора с геномом реципиентом (Wallau *et al.*, 2012). Поскольку ГП *CR1B* non-LTR ретротранспозонов и *MLE* ДНК транспозонов произошел между представителями отряда *Lepidoptera*, можно предположить, что в роли переносчика мог выступать любой общий агент: бактерия, внутриклеточный паразит, симбионт, вирус и т.д.

В настоящее время относительно детально изучены паразиты тутового шелкопряда *B. mori*. В список основных патогенов данного организма входят: бактерии (*Bacillus thuringiensis* и *Bacillus bombysepticus* (Firmicutes)), грибы (*Metarhizium anisopliae* и *Beauveria bassiana* (Ascomycota)), микроспоридии (*Nosema bombycis* и *Vairimorpha bombycis* SD) и вирусы (*Bombyx mori* nucleopolyhedrovirus T3, *Bombyx mandarina* nucleopolyhedrovirus S1, *Bombyx mori* cyrovirus 1 Suzhou, Macula-like virus) (согласно SilkPathDB: <http://silkpathdb.swu.edu.cn/silkpathdb/>). Информация о паразитах представителей рода *Maculinea* полностью отсутствует в литературе.

Вышеописанные паразиты тесно взаимодействуют с геномом хозяина, не обладают специфичностью по отношению к хозяину и могут являться потенциальными векторами ГП МГЭ. Так, для паразита *Nosema bombycis* показан случай ГП ДНК транспозонов *hAT* и MITE между геномом *Nosema bombycis* и геномом *Bombyx mori* (Zhang *et al.*, 2013a; Zhang *et al.*, 2013b).

Одна из стадий жизненного цикла (стадия гусеницы) бабочек *Maculinea* протекает в муравьином гнезде муравьев рода *Myrmica*. Муравьи кормят личинок

бабочек также как и своих собственных личинок. Гусеницы проводят в муравейнике достаточно длительный период времени вплоть до превращения в имаго. Моли семейства *Bombycidae* могут попасть в муравейник в качестве объекта питания. Таким образом, муравейник представителей рода *Myrmica* является идеальным местом для прямого контакта бабочек *Maculinea* и молей семейства *Bombycidae* и перекрестного заражения паразитами, являющимися потенциальными векторами ГП.

ЗАКЛЮЧЕНИЕ

Представленная работа направлена на изучение разнообразия, распространения и эволюции мобильных генетических элементов групп *mariner*-like TIR ДНК транспозонов и *CRI* non-LTR ретротранспозонов в геномах насекомых отряда Lepidoptera.

Исследование *CRI* non-LTR ретротранспозонов в геномах 60 видов насекомых отряда Lepidoptera выявило две филогенетически группы элементов данной клады: *Aurivillius* и *Fabre*. Элементы группы *Fabre* широко распространены в геномах изученных видов отряда. Для данной группы элементов предпочтительным является вертикальное наследование и сохранение в геноме копий, характеризующихся низким разнообразием. Исключения составляют элементы группы *CRIB*, распространение которых ограничивается двумя семействами: бабочек Lycaenidae (род *Maculinea*) и молей Bombycidae (род *Bombyx* и *Oberthueria*). Неравномерно распространение *CRIB* элементов было подтверждено методом ПЦР амплификации и дот-блот гибридизации. Сравнение последовательностей данных элементов из геномов представителей семейств Bombycidae и Lycaenidae показало высокое сходство от 95.2 до 96.4%. Вышеописанные факторы являются свидетельством ГП *CRIB* элементов между геномами представителей семейств Bombycidae и Lycaenidae. Данный факт был подтвержден с помощью филогенетических и параметрических методов анализа.

Механизм ГП *CRIB* non-LTR ретротранспозонов не известен. В рамках данной работы была проведена проверка одной из гипотез, согласно которой в ГП *CRIB* принимают участие ДНК транспозоны. В геноме *Bombyx mori* были выявлены химерные конструкции, содержащие встройки *CRIB* элементов в последовательности *MLE* TIR ДНК транспозонов *Bmmar1* и *Bmmar6*.

Анализ разнообразия *MLE* TIR ДНК транспозонов представленных в генетических базах данных показал, что в геномах представителей класса Insecta представлено 13 подсемейств *mariner* и по одному подсемейству семейств *ludens* и *mori*. Разнообразие *MLE* элементов в геномах представителей отряда Lepidoptera ограничивается шестью подсемействами семейства *mariner* (*DTTMarCRI*, *cecropia*, *mauritiana*, *mellifera*, *vertumana*, *irritans*) и семейством *mori*. Экспериментальный

поиск подтвердил присутствие элементов групп *cecropia*, *mauritiana*, *mellifera*, *vertumana* и *mori* в геномах 60 изученных видов отряда Lepidoptera. В геноме *B. mori* выявлен новый элемент *BmmarY* группы *vertumana*. Элементы кластеров *cecropia*, *mellifera*, *mauritiana*, *irritans* и *DTTMarCRI*, по-видимому, эволюционировали в геномах представителей отряда Lepidoptera путем вертикального наследования.

Элементы *BmmarY-like* (*vertumana*) и *BmmarI-like* (*mori*) представлены исключительно в геномах представителей семейств Lycaenidae (род *Maculinea*) и Bombycidae (род *Bombyx*) и характеризуются высоким сходством последовательностей между эволюционно удаленными видами. ГП элементов *BmmarY-like* и *BmmarI-like* между бабочками Lycaenidae и молями Bombycidae был подтвержден филогенетическими и параметрическими методами. В геномах представителей Lycaenidae (род *Maculinea*) были выявлены химерные конструкторы, содержащие *CRIB* элементы и один из *MLE* ДНК транспозонов (*BmmarY-like* и *BmmarI-like*).

Процессы ГП *CRIB* non-LTR ретротранспозонов и *MLE* ДНК транспозонов между бабочками Lycaenidae и молями Bombycidae могут являться независимыми или произошли совместно. В пользу совместного ГП и участия *MLE* ДНК транспозонов в ГП *CRIB* non-LTR ретротранспозонов свидетельствует наличие химерных конструкций.

Таким образом, в результате настоящего исследования было изучено разнообразие, распространение и эволюция МГЭ групп *mariner-like* TIR ДНК транспозонов и *CR1* non-LTR ретротранспозонов в геномах насекомых отряда Lepidoptera. Были выявлены случаи ГП *CRIB* non-LTR ретротранспозонов и *BmmarY-like* и *BmmarI-like* *MLE* ДНК транспозонов между бабочками Lycaenidae и молями Bombycidae. Получены косвенные подтверждения участия *MLE* ДНК транспозонов в ГП *CRIB* non-LTR ретротранспозонов. Перспективным направлением дальнейшего развития данной темы исследования является поиск векторов ГП и выявление механизмов ГП элементов групп *CRIB* и *MLE*.

ВЫВОДЫ

1. Проведен поиск non-LTR ретротранспозонов в геномах 60 видов отряда Lepidoptera и установлены их филогенетические взаимоотношения. Показано присутствие четырех филогенетических групп non-LTR ретротранспозонов (*CRI*, *TIQ*, *Jockey* и *RI*) в геномах представителей отряда Lepidoptera. Элементы группы *CRI* представлены двумя кластерами *Aurivillius* и *Fabre*.
2. Подтвержден факт горизонтального переноса *CRIB* non-LTR ретротранспозонов между геномами представителей семейств Lycaenidae (род *Maculinea*) и Bombycidae (рода *Bombyx* и *Oberthueria*).
3. В геномах видов отряда Lepidoptera выявлено шесть подсемейств семейства *mariner* (*DTTMarCRI*, *cecropia*, *mauritiana*, *mellifera*, *vertumana*, *irritans*) и семейство *mori mariner-like* ДНК транспозонов. Установлены их филогенетические взаимоотношения.
4. В геноме *Bombyx mori* обнаружен новый элемент *BmmarY*, принадлежащий к подсемейству *vertumana mariner-like* ДНК транспозонов.
5. Установлен факт горизонтального переноса *BmmarY-like* и *Bmmar1-like* ДНК транспозонов между геномами представителей семейств Lycaenidae (род *Maculinea*) и Bombycidae (род *Bombyx*).
6. В геномах *Bombyx mori*, *Maculinea teleius* и *Maculinea arion* были выявлены химерные *mariner-like* ДНК транспозоны, содержащие встройки *CRIB* non-LTR ретротранспозонов. Данные конструкции могут являться результатом совместного горизонтального переноса *CRIB* non-LTR ретротранспозонов и *mariner-like* ДНК транспозонов, что свидетельствует об участии *mariner-like* элементов в горизонтальном переносе *CRIB* элементов между видами отряда Lepidoptera.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

а.о. - аминокислотных остатков

ГП - горизонтальный перенос

е.а. - единица активности

МГЭ - мобильные генетические элементы

ПЦР - полимеразная цепная реакция

п.о. - пара оснований

5' UTR и 3' UTR - 5' и 3' нетранслируемые области

АР- эндонуклеаза (APE) - апурин/апиримидин эндонуклеаза

АТР - ATPase - АТФаза

BD - C-terminal nucleic acid binding domain - C - концевой домен связывающий нуклеиновые кислот

bp - base pairs - пар нуклеотидов

BS - binding site - сайт связывания

C-INT - C-integrase - интегразы типа C

COI - ген первой субъединицы митохондриальной цитохромоксидазы

COII - ген второй субъединицы митохондриальной цитохромоксидазы

CUB - Codon usage bias - частота встречаемости синонимичных кодонов

CYP - Cysteine protease - цистеиновая протеаза

DIR - dinucleotide inverted repeat - короткие инвертированные повторы

DIRS - Dictyostelium intermediate repeat sequence - переходные повторяющиеся последовательности *Dictyostelium*

dNTP - дезоксинуклеотид трифосфаты

EF1 α - elongation factor 1 alpha - ген фактора элонгации

EN - endonuclease domain - домен эндонуклеазы

ENV - Envelope protein - белок оболочки

GRPR - аминокислотная последовательность глицин-аргинин-пролин-аргинин

H3 - histone H3 gene - ген гистона H3

HEL - DNA helicase - ДНК геликаза

HMM - Hidden Markov Model - алгоритм основанный на скрытых цепях Маркова

HTH - helix - turn - helix - мотив спираль-поворот-спираль

HYR - tyrosine recombinase - тирозин рекомбиназа

INT - integrase - интеграз

IRES - Internal Ribosomal Entry Site - участок внутренней посадки рибосомы

LINE - Long Interspersed Nuclear Element - длинный диспергированный ядерный повтор, non-LTR ретротранспозон.

LTR - long terminal repeat – длинный концевой повтор

LZ - leucine zipper domain - домен лейцинового zipper («молнии»)

MITEs - miniature inverted repeat transposable elements - миниатюрные транспозирующие элементы, содержащие инвертированные повторы

ML - Maximum Likelihood - метод максимального правдоподобия

MLE - *Mariner*-like element - *mariner* - подобный элемент (*MLE* элемент)

MLEs - *Mariner*-like elements - *mariner*-like подобные элементы (*MLE* элементы)

MP - Maximum Parsimony - метод максимальной экономии

Mya - million years ago - миллионы лет назад

NJ - Neighbor Joining - метод соединения ближайших соседей

NLS - nuclear location signal - сигнал ядерной локализации

NS - non-significant - не достоверно

OPI - overproduction inhibition - ингибирование гиперпродукции

ORF - open reading frame - открытая рамка считывания

ORFp - белок, кодируемый ORF

PBS - primer-binding site - праймер-связывающий сайт

PEC - Paired-End Complex

PLE - Penelope-like elements - Penelope-подобные элементы

POL B - DNA polymerase B (ДНК полимеразы B),

PR - protease - протеаза

REL-endo (EN) - restriction enzyme-like endonuclease - эндонуклеаза рестриктазного типа

RH - RNase H (RNH) - рибонуклеаза H

RNPs - ribonucleoprotein particles - рибонуклеопротеидные комплексы

RPA - replication protein A - белок инициации репликации

RT - reverse transcriptase - обратная транскриптаза

SEC - Single-End Complex

SINE - Short Interspersed Nuclear Element - короткий диспергированный ядерный повтор.

TCC - Target Capture Complex - комплекс захвата мишени

TIR - terminal inverted repeat - концевые инвертированные повторы

TPRT - target primed reverse transcription - РНК-опосредованная интеграция

TSD - target site duplication - дупликация сайта - мишени

UPGMA - Unweighted Pair Group Method using arithmetic Averages - алгоритм попарной группировки

Wg - wingless ген

WGS - whole genome sequencing - полногеномное секвенирование

СПИСОК ЛИТЕРАТУРЫ

1. Aksoy S., Williams S., Chang S., Richards, F.F. SLACS Retrotransposon from *Trypanosoma brucei gambiense* Is Similar to Mammalian LINEs // *Nucleic Acid Res.* - 1990. - Vol. 18. - P. 785–792.
2. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool // *J. Mol. Biol.* - 1990. - Vol. 215. - P. 403-410.
3. Alzohairy A.M., Gyulai G., Jansen R.K., Bahieldin A. Transposable elements domesticated and neofunctionalized by eukaryotic genomes // *Plasmid.* - 2013. - Vol. 69. - P. 1-15.
4. Anzai T., Osanai M., Hamada M., Fujiwara H. Functional roles of 3'-terminal structures of template RNA during *in vivo* retrotransposition of non-LTR retrotransposon, R1Bm // *Nucleic Acids Res.* - 2005. - Vol. 33. - P. 1993-2002.
5. Aparicio S. *et al.* Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes* // *Science.* - 2002. - Vol. 297. - P. 1301-1310.
6. Arkhipova I.R., Meselson M. Diverse DNA transposons in rotifers of the class Bdelloidea // *Proc. Natl. Acad. Sci. U S A.* - 2005. - Vol. 102. - P. 11781-11786.
7. Arkhipova I.R., Morrison H.G. Three Retrotransposon Families in the Genome of *Giardia lamblia*: Two Telomeric, One Dead // *Proc. Natl. Acad. Sci. USA.* - 2001. - Vol. 98. - P. 14497–14502.
8. Aronovich E.L., Bell J.B., Belur L.R., Gunther R., Koniar B., Erickson D.C., Schachern P.A., Matisse I., McIvor R.S., Whitley C.B., Hackett P.B. Prolonged expression of a lysosomal enzyme in mouse liver after Sleeping Beauty transposon-mediated gene delivery: implications for non-viral gene therapy of mucopolysaccharidoses // *J Gene Med.* - 2007. - Vol. 9. - P. 403–415.
9. Arvestad L. Efficient methods for estimating amino acid replacement rates // *J. Mol. Evol.* - 2006. - Vol. 62. - P. 663–673.
10. Babushok D.V., Kazazian H.H. Progress in Understanding the Biology of the Human Mutagen LINE-1 // *Hum. Mutat.* - 2007. - Vol. 28. - P. 527–539.

11. Bartolome C., Bello X., Maside X. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes // *Genome Biol.* - 2009. - Vol. 10. - P. R22.
12. Bellofatto V., Cooper R., Cross G.A.M. Discontinuous Transcription in *Leptomonas seymouri*: Presence of Intact and Interrupted Mini-Exon Gene Families // *Nucleic Acids Res.* - 1988. - Vol. 16. - P. 7437–7456.
13. Berezikov E., Bucheton A., Busseau I. A search for reverse transcriptase-coding sequences reveals new non-LTR retrotransposons in the genome of *Drosophila melanogaster* // *Genome Biol.* - 2000. - Vol. 1. - P. RESEARCH0012.
14. Bergman C.M., Quesneville H. Discovering and detecting transposable elements in genome sequences // *Brief Bioinform.* - 2007. - Vol. 8. - P. 382-392.
15. Biedler J., Tu Z. Non-LTR retrotransposons in the african malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity. *Mol. Biol. Evol.* - 2003. - Vol. 20. - P. 1811–1825.
16. Biemont C., Vieira C. Genetics: junk DNA as an evolutionary force // *Nature.* - 2006. - Vol. 443. - P. 521–524.
17. Blinov A.G., Sobanov Y.V., Scherbik S.V., Aimanova K.G. The *Chironomus* (*Camptochironomus*) *tentans* genome contains two non-LTR retrotransposons // *Genome.* - 1997. - Vol. 40. - P. 143–150.
18. Blumenstiel J.P. Evolutionary dynamics of transposable elements in a small RNA world // *Trends Genet.* - 2011. - Vol. 27. - P. 23-31.
19. Brilllet B., Bigot Y., Auge-Gouillou C. Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains // *Genetica.* - 2007. - Vol. 130. - P. 105-120.
20. Bringaud F., Ghedin E., Blandin G., Bartholomeu D.C., Caler E., Levin M.J., Baltz T., El-Sayed N.M. Evolution of Non-LTR Retrotransposons in the Trypanosomatid Genomes: *Leishmania major* Has Lost the Active Elements // *Mol. Biochem. Parasitol.* - 2006. - Vol. 145. - P. 158–170.
21. Bromham L., Penny D. The modern molecular clock // *Nat Rev Genet.* - 2003. - Vol. 4. - P. 216-224.

22. Bowers W.J., Mastrangelo M.A., Howard D.F., Southerland H.A., Maguire-Zeiss K.A., Federoff H.J. Neuronal precursor-restricted transduction via in utero CNS gene delivery of a novel bipartite HSV amplicon/transposase hybrid vector // *Mol Ther.* - 2006. - Vol. 13. - P. 580–588.
23. Brouha B., Schustak J., Badge R.M., Lutz-Prigge S., Farley A.H., Moran J.V., Kazazian H.H. Jr. Hot L1s account for the bulk of retrotransposition in the human population // *Proc Natl Acad Sci USA.* - 2003. - Vol. 100. - P. 5280-5285.
24. Burch J.B., Davis D.L., Haas N.B. Chicken Repeat 1 Elements Contain a pol-Like Open Reading Frame and Belong to the Non-Long Terminal Repeat Class of Retrotransposons, *Proc. Natl. Acad. Sci. USA.* - 1993. - Vol. 90. - P. 8199–8203.
25. Burke W.D., Muller F., Eickbush T.H. R4, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes // *Nucleic Acids Res.* - 1995. - Vol. 23. - P. 4628-4634.
26. Chambeyron S., Bucheton A., Busseau I. Tandem UAA repeats at the 3'- end of the transcript are essential for the precise initiation of reverse transcription of the I factor in *Drosophila melanogaster* // *J Biol Chem.* - 2002. - Vol. 277. - P. 17877-17882.
27. Chenais B., Caruso A., Hiard S., Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments // *Gene.* - 2012. - Vol. 509. - P. 7-15.
28. Chen J.M., Stenson P.D., Cooper D.N., Ferec C. 2005. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease // *Hum. Genet.* - 2005. - Vol. 117. - P. 411–427.
29. Chen Y., Zhou F., Li G., Xu Y. MUST: a system for identification of miniature inverted-repeat transposable elements and applications to *Anabaena variabilis* and *Haloquadratum walsbyi* // *Gene.* - 2009. - Vol. 436. - P. 1–7.
30. Cost G. J., Feng Q., Jacquier A., Boeke J. D. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* - 2002. - Vol. 21. - P. 5899–5910.
31. Christensen S.M., Eickbush T.H. R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA // *Mol Cell Biol.* - 2005. - Vol. 25. - P. 6617-6628.

32. Curcio M., Derbyshire K. The outs and ins of transposition: from mu to kangaroo // *Nature Rev. Mol.Cell Biol.* - 2003. - Vol. 4. - P. 865–877.
33. Daniels S.B., Strausbaugh L.D., Ehrman L., Armstrong R. Sequences homologous to P elements occur in *Drosophila paulistorum*. *Proc Natl Acad Sci U S A.* - 1984. - Vol. 81. - P. 6794–6797.
34. Darriba D., Taboada G.L., Doallo R., Posada D. ProtTest 3: fast selection of best-fit models of protein evolution // *Bioinformatics.* 2011. – Vol. 27. - P. 1164-1165.
35. Darriba D., Taboada G.L., Doallo R., Posada D. jModelTest 2: more models, new heuristics and parallel computing // *Nature Methods.* - 2012. - Vol. 9. - P. 772.
36. Dawson A., Hartswood E., Paterson T., Finnegan D.J. A LINE-like transposable element in *Drosophila*, the I factor, encodes a protein with properties similar to those of retroviral nucleocapsids // *EMBO J.* - 1997. - Vol. 16. - P. 4448-4455.
37. DeMarco R., Venancio T.M., Verjovski-Almeida S. SmTRC1, a novel *Schistosoma mansoni* DNA transposon, discloses new families of animal and fungi transposons belonging to the CACTA superfamily // *BMC Evol Biol.* - 2006. - Vol. 6. - P. 89.
38. Di Matteo M., Matrai J., Belay E., Firdissa T., Vandendriessche T., Chuah M.K. PiggyBac toolbox // *Methods Mol Biol.* - 2012. - Vol. 859. - P. 241-254.
39. Dlakic M. Functionally unrelated signalling proteins contain a fold similar to Mg²⁺-dependent endonucleases // *Trends Biochem Sci.* - 2000. - Vol. 25. - P. 272–273.
40. Douzery E.J., Snell E.A., Baptiste E., Delsuc F., Philippe H. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc Natl Acad Sci USA.* - 2004. - Vol. 101. - P. 15386–15391.
41. Du C., Caronna J., He L., Dooner H.K. Computational prediction and molecular confirmation of Helitron transposons in the maize genome // *BMC Genomics.* - 2008. - Vol. 9. - P. 51.
42. Duan J., Li R., Cheng D., Fan W., Zha X., Cheng T., Wu Y., Wang J., Mita K., Xiang Z., Xia Q. SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology // *Nucleic Acids Res.* - 2010. - Vol. 38(Database issue):D453-6.

43. Dufraigne C., Fertil B., Lespinats S., Giron A., Deschavanne P. Detection and characterization of horizontal transfers in prokaryotes using genomic signature // *Nucleic Acids Res.* - 2005. - Vol. 33. - P. e6.
44. Dunning Hotopp JC, Clark ME, Oliveira DC, Foster JM, Fischer P, Muñoz Torres MC, Giebel JD, Kumar N, Ishmael N, Wang S, Ingram J, Nene RV, Shepard J, Tomkins J, Richards S, Spiro DJ, Ghedin E, Slatko BE, Tettelin H, Werren JH. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes // *Science.* - 2007. - Vol. 317. - P. 1753-1756.
45. Eddy S.R. Profile hidden Markov models // *Bioinformatics.* - 1998. - Vol. 14. - P. 755-763.
46. Eddy S.R. Accelerated Profile HMM Searches // *PLoS Comput Biol.* - 2011. - Vol. 7. - P. e1002195.
47. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput // *Nucleic Acids Res.* - 2004. - Vol. - 32. - P. 1792-1797.
48. Eickbush T.H. R2 and related site-specific non-long terminal repeat retrotransposons // T.H. Eickbush. *Mobile DNA II.* - ASM Press. 2002. - P. 813–835.
49. Eickbush T.H., Burke W.D., Eickbush D.G., Lathe W.C. 3rd. Evolution of R1 and R2 in the rDNA units of the genus *Drosophila* // *Genetica.* - 1997. - Vol. 100. - P. 49-61.
50. Eickbush T.H., Jamburuthugoda V.K. The diversity of retrotransposons and the properties of their reverse transcriptases // *Virus Res.* - 2008. - Vol. 134. - P. 221-234.
51. Eickbush T.H. Origins and Evolution of Retrotransposons // T.H. Eickbushand, H.S Malik. *Mobile DNA II.* - ASM Press. 2002. - P. 1111–1144.
52. Elhai J., Liu H., Taton A. Detection of horizontal transfer of individual genes by anomalous oligomer frequencies // *BMC Genomics.* - 2012. - Vol. 13. - P. 245.
53. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap // *Evolution.* - 1985. - Vol. 39. - P. 783-791.
54. Felsenstein J. PHYLIP - Phylogeny Inference Package (Version 3.2) // *Cladistics.* - 1989. - Vol. 5. - P. 164-166.

55. Feng Q., Moran J., Kazazian H.H.Jr., Boeke J.D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition // *Cell*. - 1996. - Vol. 87. - P. 905–916.
56. Feschotte C., Jiang N., Wessler S. R. Plant transposable elements: where genetics meets genomics // *Nat. Rev. Genet.* - 2002. - Vol. 3. - P. 329–341.
57. Feschotte C., Pritham E.J. DNA transposons and the evolution of eukaryotic genomes // *Annu Rev Genet.* - 2007. - Vol. 41. - P. 331-368.
58. Feschotte C., Wessler S.R. Mariner-like transposases are widespread and diverse in flowering plants // *Proc Natl Acad Sci USA*. - 2002. - Vol. 99. - P. 280-285.
59. Feschotte C., Swamy L., Wessler S.R. Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with Stowaway MITEs // *Genetics*. - 2003. - Vol. 163. - P. 747–758.
60. Fillingham J.S., Thing T.A., Vythilingum N., Keuroghlian A., Bruno D., Golding G.B., Pearlman R.E. A Non-Long Terminal Repeat Retrotransposon Family Is Restricted to the Germ Line Micronucleus of the Ciliated Protozoan *Tetrahymena thermophila* // *Eukaryot. Cell*. - 2004. - Vol. 3. - P. 157–169.
61. Finnegan D.J. Eukaryotic transposable elements and genome evolution // *Trends Genet.* - 1989. - Vol. 5. - P. 103–107.
62. Friesen P.D., Nissen M.S. Gene organization and transcription of ted, a lepidopteran retrotransposon integrated within the baculovirus genome // *Mol. Cell. Biol.* - 1990. - Vol. 10. - P. 3067–3077.
63. Gaunt M.W., Miles M.A. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks // *Mol Biol Evol.* - 2002. - Vol. 19. - P. 748–761.
64. George J.A., Eickbush T.H. Conserved features at the 5' end of *Drosophila* R2 retrotransposable elements: implications for transcription and translation // *Insect Mol Biol.* - 1999. - Vol. 8. - P. 3-10.
65. Gilbert C., Schaack S., Pace J.K. 2nd, Brindley P.J., Feschotte C. A role for host-parasite interactions in the horizontal transfer of transposons across phyla // *Nature*. - 2010. - Vol. 464. - P. 1347-1350.

66. Goodwin T.J., Ormandy J.E., Poulter R.T. L1- Like Non-LTR Retrotransposons in the Yeast *Candida albicans* // *Curr. Genet.* - 2001. - Vol. 39. - P. 83–91.
67. Goodwin T.J., Poulter R.T. The diversity of retrotransposons in the yeast *Cryptococcus neoformans* // *Yeast.* - 2001. - Vol. 18. - P. 865–880.
68. Gu W., Castoe T.A., Hedges D.J., Batzer M.A., Pollock D.D. Identification of repeat structure in large genomes using repeat probability clouds // *Anal Biochem.* - 2008. - Vol. 380. - P. 77-83.
69. Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0 // *Syst Biol.* - 2010. - Vol. 59. - P. 307-321.
70. Haas B.J., *et al.* Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature.* - 2009. - Vol. 461. - P. 393–398.
71. Han J.S. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions // *Mobile DNA.* - 2010. - Vol. 1. - P. 15.
72. Han J.S., Boeke J.D. A highly active synthetic mammalian retrotransposon // *Nature.* - 2004. - Vol. 429. - P. 314-318.
73. Hartl D.L., Lohe A.R., Lozovskaya E.R. Modern thoughts on an ancient mariner: function, evolution, regulation // *Annu Rev Genet.* - 1997. - Vol. 31. - P. 337-358.
74. Hecht M.M., Nitz N., Araujo P.F., Sousa A.O., Rosa Ade C., Gomes D.A., Leonardecz E., Teixeira A.R. Inheritance of DNA transferred from American trypanosomes to human hosts // *PLoS One.* - 2010. - Vol. 5. - P. e9181.
75. Hickman A.B., Perez Z.N., Zhou L., Musingarimi P., Ghirlando R., Hinshaw J.E., Craig N.L., Dyda F. Molecular architecture of a eukaryotic transposase // *Nat. Struct. Mol. Biol.* - 2005. - Vol. 12. - P. 715–721.
76. Hofmann K., Tomiuk S., Wolff G., Stoffel W. Cloning and characterization of the mammalian brain-specific, Mg²⁺ -dependent neutral sphingomyelinase // *Proc Natl Acad Sci USA.* - 2000. - Vol. 97. - P. 5895–5900.

77. Hohjoh H., M. F. Singer. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon // *EMBO J.* - 1997. - Vol. 16. - P. 6034–6043.
78. Houck M.A., Clark J.B., Peterson K.R., Kidwell M.G. Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis* // *Science.* - 1991. - Vol. 253. - P. 1125-1128.
79. Huelsenbeck J.P., Ronquist F. MrBayes: Bayesian inference of phylogenetic trees // *Bioinformatics.* - 2001. - Vol. 17. - P. 754–755
80. Ivics Z., Izsvak Z., Minter A., Hackett P.B. Identification of functional domains and evolution of Tc1-like transposable elements // *Proc Natl Acad Sci USA.* - 1996. - Vol. 93. - P. 5008-5013.
81. Izsvak Z., Chuah M.K., Vandendriessche T., Ivics Z. Efficient stable gene transfer into human cells by the Sleeping Beauty transposon vectors // *Methods.* - 2009. - Vol. 49. - P. 287-297.
82. Jeffs A.R., Benjes S.M., Smith T.L., Sowerby S.J., Morris C.M. The BCR gene recombines preferentially with Alu elements in complex BCR-ABL translocations of chronic myeloid leukaemia. *Hum. Mol. Genet.* - 1998. - Vol. 7. - P. 767–776.
83. Jehle J.A., Nickel A., Vlak J.M., Backhaus H. Horizontal escape of the novel Tc1-like lepidopteran transposon TCp3.2 into *Cydia pomonella* granulovirus // *J Mol Evol.* - 1998. - Vol. 46. - P. 215-224.
84. Jiang N., Bao Z., Temnykh S., Cheng Z., Jiang J., Wing R.A., McCouch S.R., Wessler S.R. Dasheng: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice // *Genetics.* - 2002. - Vol. 161. - P. 1293–1305.
85. Jiang N., Bao Z., Zhang X., Hirochika H., Eddy S.R., McCouch S.R., Wessler S.R. An active DNA transposon family in rice // *Nature.* - 2003. - Vol. 421. - P. 163-167.
86. Jordan I.K., Bowen N.J. Computational analysis of transposable element sequences // *Methods Mol Biol.* - 2004. - Vol. 260. - P. 59-71.

87. Jordan I.K., Rogozin I.B., Glazko G.V., Koonin E.V. Origin of a substantial fraction of human regulatory sequences from transposable elements // *Trends Genet.* - 2003. - Vol. 19. - P. 68–72.
88. Kaer K., Speek M. Retroelements in human disease // *Gene.* - 2013. - Vol. 518. - P. 231-241.
89. Kajikawa M., Okada N. LINEs mobilize SINEs in the eel through a shared 3' sequence // *Cell.* - 2002. - Vol. 111. - P. 433–444.
90. Kalafus K.J., Jackson A.R., Milosavljevic A. Pash: efficient genome-scale sequence anchoring by Positional Hashing // *Genome Res.* - 2004. - Vol. 14. - P. 672-678.
91. Kalendar R., Flavell A.J., Ellis T.H., Sjakste T., Moisy C., Schulman A.H. Analysis of plant diversity with retrotransposon-based molecular markers // *Heredity (Edinb).* - 2011. - Vol.106. - P. 520-530.
92. Kapitonov V.V., Jurka J. The esterase and PHD domains in CR1-like non-LTR retrotransposons // *Mol Biol Evol.* - 2003. - Vol. 20. - P. 38-46.
93. Kent W.J. BLAT-the BLAST-like alignment tool // *Genome Research.* - 2002. - Vol. 12. - P. 656–664.
94. Kim A., Terzian C., Santamaria P., Pelisson A., Purd'homme N., Bucheton A. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster* // *Proc Natl Acad Sci USA.* - 1994. - Vol. 91. - P. 1285–1289.
95. Kirilyuk A., Tolstonog G.V., Damert A., Held U., Hahn S., Lower R., Buschmann C., Horn A.V., Traub P., Schumann G.G. Functional endogenous LINE-1 retrotransposons are expressed and mobilized in rat chloroleukemia cells // *Nucleic Acids Res.* - 2008. - Vol. 36. - P. 648-665.
96. Kojima K.K., Fujiwara H. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets // *Mol Biol Evol.* - 2004. - Vol. 21. - P. 207-217.
97. Kojima K.K., Fujiwara H. An Extraordinary Retrotransposon Family Encoding Dual Endonucleases // *Genome Res.* - 2005. - Vol. 15. - P. 1106–1117.

98. Kolosha V.O., Martin S.L. High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1) // *J Biol Chem.* - 2003. - Vol. 278. - P. 8112-8117.
99. Kordis D., Gubensek F. The Bov-B lines found in *Vipera ammodytes* toxic PLA2 genes are widespread in snake genomes // *Toxicon.* - 1998. - Vol. 36. - P. 1585-1590.
100. Kristensen, N.P. *Lepidoptera, Moths and Butterflies. Vol. 2: Morphology, Physiology, and Development* / N.P. Kristensen. - Berlin & New York: Walter de Gruyter, 2003. 564 pp.
101. Kumaresan G., Mathavan S. Molecular diversity and phylogenetic analysis of mariner-like transposons in the genome of the silkworm *Bombyx mori* // *Insect Mol Biol.* - 2004. - Vol. 13. - P. 259-271.
102. Kurtz S., Choudhuri J.V., Ohlebusch E., Schleiermacher C., Stoye J., Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale // *Nucleic Acids Res.* - 2001. - Vol. 29. - P. 4633-4642.
103. Lampe D.J., Grant T.E., Robertson H.M. Factors affecting transposition of the Himar1 mariner transposon in vitro // *Genetics.* - 1998. - Vol. 149. - P. 179-187.
104. Lampe D.J., Witherspoon D.J., Soto-Adames F.N., Robertson H.M. Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer // *Mol Biol Evol.* - 2003. - Vol. 20. - P 554-562.
105. Lander E.S. *et al.* Initial sequencing and analysis of the human genome // *Nature.* - 2001. - Vol. 409. - P. 860-921.
106. Leroy H., Castagnone-Sereno P., Renault S., Auge-Gouillou C., Bigot Y., Abad P. Characterization of Mymar1, a mariner-like element with large inverted terminal repeats (ITRs) from the phytoparasitic nematode *Meloidogyne chitwoodi* // *Gene.* - 2003. - Vol. 304. - P. 35-41.
107. Li R., Ye J., Li S., Wang J., Han Y., Ye C., Wang J., Yang H., Yu J., Wong G.K., Wang J. ReAS: recovery of ancestral sequences for transposable elements from the unassembled reads of a whole genome shotgun // *PLoS Comput. Biol.* - 2005. - Vol. 1. - P. e43.

108. Lim J.K., Simmons M.J. Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster* // *Bioessays*. - 1994. - Vol. 16. - P. 269–275.
109. Liu D., Bischerour J., Siddique A., Buisine N., Bigot Y., Chalmers R. The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase // *Mol Cell Biol*. - 2007. - Vol. 27. - P. 1125-1132.
110. Lohe A.R., Hartl D.L. Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation // *Mol Biol Evol*. - 1996. - Vol. 13. - P. 549-555.
111. Lohe A.R., Moriyama E.N., Lidholm D.A., Hartl D.L. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements // *Mol Biol Evol*. - 1995. - Vol. 12. - P. 62–72.
112. Loreto E.L., Carareto C.M., Capy P. Revisiting horizontal transfer of transposable elements in *Drosophila* // *Heredity (Edinb)*. - 2008. - Vol. 100. - P. 545-554.
113. Lovsin N., Gubensek F., Kordis D. Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in *Deuterostomia* // *Mol. Biol. Evol*. - 2001. - Vol. 18. - P. 2213-2224.
114. Lu G., Moriyama E.N. Vector NTI, a balanced all-in-one sequence analysis suite // *Brief Bioinform*. - 2004. - Vol. 5. - P. 378-388.
115. Lucier J.F., Perreault J., Noël J.F., Boire G., Perreault J.P. RTAnalyzer: a web application for finding new retrotransposons and detecting L1 retrotransposition signatures // *Nucleic Acids Res*. - 2007. - Vol. 35. - P.W269-W274.
116. Ludwig A., Valente V.L., Loreto E.L. Multiple invasions of Errantivirus in the genus *Drosophila* // *Insect Mol Biol*. - 2008. - Vol. 17. - P.113–124.
117. Malik H.S., Burke W.D., Eickbush T.H. The age and evolution of non-LTR retrotransposable elements // *Mol Biol Evol*. - 1999. - Vol. 16. - P. 793–805.
118. Malik H.S., Eickbush T.H. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs // *Mol Biol Evol*. - 1998. - Vol. 15. - P.1123-1134.

119. Malik H.S., Eickbush H.T. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses // *Gen. Res.* - 2001. - Vol. 11. - P. 1187–1197.
120. Marin I., Plata-Rengifo P., Labrador M., Fontdevila A. Evolutionary Relationships among the Members of an Ancient Class of Non-LTR Retrotransposons Found in the Nematode *Caenorhabditis elegans* // *Mol.Biol. Evol.* - 1998. - Vol. 15. - P. 1390–1402.
121. Martin S.L., Bushman F.D. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon // *Mol Cell Biol.* - 2001. - Vol. 21. - P. 467–475.
122. Martin F., Maranon C., Olivares M., Alonso C., Lopez M.C. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the Ape family of DNA repair enzymes // *J Mol Biol.* - 1995. - Vol. 247. - P. 49–59.
123. Martinez-Abarca F., Toro N. Group II introns in the bacterial world // *Mol Microbiol.* - 2000. - Vol. 38. - P. 917-926.
124. Matsumoto T., Takahashi H., Fujiwara H. Targeted nuclear import of open reading frame 1 protein is required for in vivo retrotransposition of a telomere-specific non-long terminal repeat retrotransposon, SART1 // *Mol Cell Biol.* - 2004. - Vol. 24. - P. 105-122.
125. McCarthy E.M., McDonald J.F. LTR_STRUC: a novel search and identification program for LTR retrotransposons // *Bioinformatics.* - 2003. - Vol. 19. - P. 362-367.
126. McClintock B. Chromosome organization and genic expression // *Cold Spring Harbor Symp. Quant. Biol.* - 1951. - Vol. 16. - P. 13–47.
127. Meyers B.C., Tingey S.V., Morgante M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome // *Genome Res.* - 2001. - Vol. 11. - P. 1660–1676.
128. Moran J.V., Gilbert N. Mammalian LINE-1 retrotransposons and related elements // J.V. Moran, N. Gilbert. *Mobile DNA II.* - ASM Press. 2002. - P. 836-869.
129. Morgulis A., Gertz E.M., Schäffer A.A., Agarwala R. WindowMasker: window-based masker for sequenced genomes // *Bioinformatics.* - 2006. - Vol. 22. - P. 134-141.

130. Mota N.R., Ludwig A., Valente V.L., Loreto E.L. Harrow: new *Drosophila* hAT transposons involved in horizontal transfer // *Insect Mol Biol.* - 2010. - Vol. 19. - P. 217–228.
131. Munoz-Lopez M., Garcia-Perez J.L. DNA transposons: nature and applications in genomics // *Curr Genomics.* - 2010. - Vol. 11. - P. 115-128.
132. Nakamura T.M., Morin G.B., Chapman K.B., Weinrich S.L., Andrews W., Lingner J., Harley C.B., Cech T.R. Telomerase catalytic subunit homologs from fission yeast and human // *Science.* - 1997. - Vol. 277. - P. 955–959.
133. Nazari V., Zakharov E.V., Sperling F.A. Phylogeny, historical biogeography, and taxonomic ranking of Parnassiinae (Lepidoptera, Papilionidae) based on morphology and seven genes // *Mol Phylogenet Evol.* - 2007. - Vol. 42. - P. 131–156.
134. Ning Z., Cox A.J., Mullikin J.C. SSAHA: a fast search method for large DNA databases // *Genome Res.* - 2001. - Vol. 11. - P. 1725-1729.
135. Notredame C., Higgins D.G., Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment // *J Mol Biol.* - 2000. - Vol. 302. - P. 205-217.
136. Novikova O.S., Blinov A.G. Origin, evolution, and distribution of different groups of non-LTR retrotransposons among eukaryotes // *Genetika.* - 2009. - Vol. 45. - P.149-159.
137. Novikova O., Sliwinska E., Fet V., Settele J., Blinov A., Woyciechowski M. CR1 clade of non-LTR retrotransposons from Maculinea butterflies (Lepidoptera: Lycaenidae): evidence for recent horizontal transmission // *BMC Evol Biol.* - 2007. - Vol. 7. - P. 93.
138. Novikova O., Fet V., Blinov A. Non-LTR retrotransposons in fungi // *Genomics.* - 2009. - Vol. 9. - P. 27-42.
139. Ohlfest J.R., Lobitz P.D., Perkinson S.G., Largaespada D.A. Integration and long-term expression in xenografted human glioblastoma cells using a plasmid-based transposon system // *Mol Ther.* - 2004. - Vol.10. - P. 260–268.
140. Okonechnikov K., Golosova O., Fursov M., Varlamov A., Vaskin Y., Efremov I., Grehov G., Kandrov D., Rasputin K., Syabro M., Tleukenov T. Unipro UGENE: a unified bioinformatics toolkit // *Bioinformatics.* - 2012. - Vol. 28. - P. 1166-1167.

141. Osanai M., Takahashi H., Kojima K.K., Hamada M., Fujiwara H. Essential motifs in the 3' untranslated region required for retrotransposition and the precise start of reverse transcription in non-long-terminal-repeat retrotransposon SART1 // *Mol Cell Biol.* - 2004. - Vol. 24. - P. 7902-7913.
142. Ostertag E.M., Goodier J.L., Zhang Y., Kazazian Jr.H.H. SVA elements are nonautonomous retrotransposons that cause disease in humans // *Am. J. Hum. Genet.* - 2003. - Vol. 73. - P. 1444–1451.
143. Ott M.G., Schmidt M., Schwarzwaelder K., Stein S., Siler U., Koehl U., *et al.* Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1 // *Nat Med.* - 2006. - Vol. 12. - P. 401–409.
144. Pearson W.R., Lipman D.J. Improved tools for biological sequence comparison // *Proc. Natl. Acad. Sci. USA.* - 1988. - Vol. 85. - P. 2444-2448.
145. Pace J.K., Gilbert C., Clark M.S., Feschotte C. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods // *Proc. Natl. Acad. Sci. USA.* - 2008. - Vol. 105. - P. 17023-17028.
146. Peng P.D., Cohen C.J., Yang S., Hsu C., Jones S., Zhao Y., Zheng Z., Rosenberg S.A., Morgan R.A. Efficient nonviral Sleeping Beauty transposon-based TCR gene transfer to peripheral blood lymphocytes confers antigen-specific antitumor reactivity // *Gene Ther.* - 2009. - Vol. 16. - P. 1042–1049.
147. Permanyer J., Gonzalez-Duarte R., Albalat R. The Non-LTR Retrotransposons in *Ciona intestinalis*: New Insights into the Evolution of Chordate Genomes // *Genome Biol.* - 2003. - Vol. 4. - P. R73.
148. Pevzner P.A., Tang H., Tesler G. De novo repeat classification and fragment assembly // *Genome Res.* - 2004. Vol. 14. - P. 1786-1796.
149. Pietrokovski S., Henikoff S. A helix–turn–helix DNA-binding motif predicted for transposases of DNA transposons // *Mol. Gen. Genet.* - 1997. - Vol. 254. - P. 689–695.
150. Piskurek O., Okada N. Poxviruses as possible vectors for horizontal transfer of retrotransposons from reptiles to mammals // *Proc. Natl. Acad. Sci. USA.* - 2007. - Vol. 104. - P. 12046–12051.

151. Plasterk R.H., Izsvak Z., Ivics Z. Resident aliens: the Tc1/mariner superfamily of transposable elements // *Trends Genet.* - 1999. - Vol. 15. - P. 326-332.
152. Ren X., Park Y., Miller T.A. 2006. Intact mariner-like element in tobacco budworm, *Heliothis virescens* (Lepidoptera: Noctuidae) // *Insect Mol Biol.* - 2006. - Vol. 6. - P. 743–748.
153. Richardson J.M., Dawson A., O'Hagan N., Taylor P., Finnegan D.J., Walkinshaw M.D. Mechanism of Mos1 transposition: insights from structural analysis // *EMBO J.* - 2006. - Vol. 25. - P. 1324-1334.
154. Robertson H.M. The mariner transposable element is widespread in insects // *Nature.* - 1993. - Vol. 362. - P. 241-245.
155. Robertson H.M., Asplund M.L. Bmmar1: a basal lineage of the mariner family of transposable elements in the silkworm moth, *Bombyx mori* // *Insect Biochem Mol Biol.* - 1996. - Vol. 26. - P. 945–954.
156. Robertson H.M., Walden K.K. Bmmar6, a second mori subfamily mariner transposon from the silkworm moth *Bombyx mori* // *Insect Mol Biol.* - 2003. - Vol. 12. - P. 167–171.
157. Rostant W.G., Wedell N., Hosken D.J. Transposable elements and insecticide resistance // *Adv Genet.* - 2012. - Vol. 78. - P. 169-201.
158. Rouault J.D., Casse N., Chenais B., Hua-Van A., Filee J., Capy P. Automatic classification within families of transposable elements: application to the mariner Family // *Gene.* - 2009. - Vol. 448. - P. 227-232.
159. Routh A., Domitrovic T., Johnson J.E. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus // *Proc. Natl. Acad. Sci. USA.* - 2012. - Vol. 109. - P. 1907-1912.
160. Roy-Engel A.M., El-Sawy M., Farooq L., Odom G.L., Perepelitsa-Belancio V., Bruch H., Oyeniran O.O., Deininger P.L. Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet. Genome Res.* - 2005. - Vol. 110. - P. 365–371.
161. Sanchez-Gracia A., Maside X., Charlesworth B. High rate of horizontal transfer of transposable elements in *Drosophila* // *Trends Genet.* - 2005. - Vol. 21. - P. 200-203.

162. Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors // *Proc. Natl. Acad. Sci. USA.* - 1977. - Vol. 74. - P. 5463-5467.
163. Schaack S., Gilbert C., Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution // *Trends Ecol. Evol.* - 2010. - Vol. 25. - P. 537-546.
164. Schulman A.H., Kalendar R. Movable feast: diverse retrotransposons and the contribution to barley (*Hordeum vulgare*) genome dynamics // *Cytogenet. Genome Res.* - 2005. - Vol. 110. - P. 598–605.
165. Seleme M.C., Busseau I., Malinsky S., Bucheton A., Teninges D. Highfrequency retrotransposition of a marked I factor in *Drosophila melanogaster* correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes // *Genetics.* - 1999. - Vol. 151. - P. 761-771.
166. Sharma R., Bagchi A., Bhattacharya A., Bhattacharya S. Characterization of a Retrotransposon-Like Element from *Entamoeba histolytica* // *Mol. Biochem. Parasitol.* - 2001. - Vol. 116. - P. 45–53.
167. Shedlock A.M. Phylogenomic Investigation of CR1 LINE Diversity in Reptiles // *Syst. Biol.* - 2006. - Vol. 55. - P. 902–911.
168. Silva J.C., Kidwell M.G. Horizontal transfer and selection in the evolution of P elements // *Mol Biol Evol.* - 2000. - Vol. 17. - P. 1542–1557.
169. Silva J.C., Loreto E.L.S., Clark J.B. Factors that affect the horizontal transfer of transposable elements // *Curr. Issues Mol. Biol.* - 2004. - Vol. 6. - P. 57–71.
170. Slotkin R.K., Martienssen R. Transposable elements and the epigenetic regulation of the genome // *Nat. Rev. Genet.* - 2007. - Vol. 8. - P. 272-285.
171. Song S.U., Gerasimova T., Kurkulos M., Boeke J.D., Corces V.G. An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus // *Gene Dev.* - 1994. - Vol. 8. - P. 2046–2057.
172. St. Laurent G. 3rd, Hammell N., McCaffrey T.A. A LINE-1 component to human aging: do LINE elements exact a longevity cost for evolutionary advantage? // *Mech Ageing Dev.* - 2010. - Vol. 131. - P. 299-305.

173. Swofford D.L. 2000. PAUP*: phylogenetic analysis by parsimony, version 4. Sinauer Associates. Massachusetts: Sanderland.
174. Szak S.T., Pickeral O.K., Makalowski W., Boguski M.S., Landsman D., Boeke J.D. Molecular archeology of L1 insertions in the human genome // *Genome Biol.* - 2002. - Vol. 3. - P. research0052.
175. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods // *Mol. Biol. Evol.* - 2011. - Vol. 28. - P. 2731-2739.
176. Thompson J.D., Higgins D.G., Gibson T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice // *Nucleic Acids Res.* - 1994. - Vol. 22. - P. 4673-4680.
177. Tu Z. Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae* // *Proc Natl Acad Sci USA.* - 2001. - Vol. 98. - P. 1699–1704.
178. Tu Z., Li S., Mao C. The changing tails of a novel short interspersed element in *Aedes aegypti*: genomic evidence for slippage retrotransposition and the relationship between 3' tandem repeats and the poly(dA) tail // *Genetics.* - 2004. - Vol. 168. - P. 2037–2047.
179. van Nieukerken E.J. Order Lepidoptera Linnaeus, 1758 // E.J. van Nieukerken, L. Kaila, I.J. Kitching, N.P. Kristensen, D.C. Lees, *et al.* *Animal Biodiversity: An outline of higher level classification and survey of taxonomic richness. Order Lepidoptera Linnaeus, 1758.* – *Zootaxa.* 2011. - P. 212–221.
180. Venner S., Feschotte C., Biemont C. Dynamics of transposable elements: towards a community ecology of the genome // *Trends Genet.* - 2009. - Vol. 25. - P. 317–323.
181. Villanueva M.S., Williams S.P., Beard C.B., Richards F.F. A New Member of a Family of Site-Specific Retrotransposons Is Present in the Spliced Leader RNA Genes of *Trypanosoma cruzi* // *Mol. Cell. Biol.* - 1991. - Vol. 11. - P. 6139–6148.

182. Volff J.N., Korting C., Froschauer A., Sweeney K., Scharl M. Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates // *J. Mol. Evol.* - 2001. - Vol. 52. - P. 351-360.
183. Volff J.N., Korting C., Scharl M. Multiple Lineages of the Non-LTR Retrotransposon Rex1 with Varying Success in Invading Fish Genomes // *Mol. Biol. Evol.* - 2000. - Vol. 17. - P. 1673–1684.
184. Volfovsky N., Haas B.J., Salzberg S.L. A clustering method for repeat analysis in DNA sequences // *Genome Biol.* - 2001. - Vol. 2. - P. RESEARCH0027.
185. Wallau G.L., Ortiz M.F., Loreto E.L. Horizontal transposon transfer in eukarya: detection, bias, and perspectives // *Genome Biol Evol.* - 2012. - Vol. 4. - P. 689-699.
186. Wang H., Hartwood E., Finnegan D.J. Pogo transposase contains a putative helix-turn-helix DNA binding domain that recognises a 12 bp sequence within the terminal inverted repeats // *Nucleic Acids Res.* - 1999. - Vol. 27. - P. 455-461.
187. Weichenrieder O., Repanas K., Perrakis A. Crystal structure of the targeting endonuclease of the human LINE-1 retrotransposon // *Structure.* - 2004. - Vol. 12. - P. 975– 986.
188. 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. A unified classification system for eukaryotic transposable elements // *Nat Rev Genet.* - 2007. - Vol. 8. - P. 973-982.
189. Witherspoon D.J., Robertson H.M. Neutral evolution of ten types of mariner transposons in the genomes of *Caenorhabditis elegans* and *C. briggsae*. *J. Mol. Evol.* - 2003. - Vol. 56. - P. 751–769.
190. Xiong Y., Eickbush T.H. Dong, a non-long terminal repeat (non-LTR) retrotransposable element from *Bombyx mori* // *Nucleic Acids Res.* - 1993. - Vol. 21. - P. 1318.
191. Xu Z., Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons // *Nucleic Acids Res.* - 2007. - Vol. 35(Web Server issue). - P. W265-W268.

192. Yang G., Hall T.C. MAK, a computational tool kit for automated MITE analysis // *Nucleic Acids Res.* - 2003. - Vol. 31. - P. 3659-3665.
193. Yang J., Malik H.S., Eickbush T.H. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements // *Proc. Natl. Acad. Sci. USA.* - 1999. - Vol. 96. - P. 7847–7852.
194. Yuan Y.W., Wessler S.R. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies // *Proc. Natl. Acad. Sci. USA.* - 2011. - Vol. 108. - P. 7884-7889.
195. Zhang H.H., Shen Y.H., Xu H.E., Liang H.Y., Han M.J., Zhang Z. A novel hAT element in *Bombyx mori* and *Rhodnius prolixus*: its relationship with miniature inverted repeat transposable elements (MITEs) and horizontal transfer // *Insect Mol. Biol.* - 2013a. - Vol. 22. - P. 584-596.
196. Zhang H.H., Xu H.E., Shen Y.H., Han M.J., Zhang Z. The origin and evolution of six miniature inverted-repeat transposable elements (MITEs) in *Bombyx mori* and *Rhodnius prolixus* // *Genome Biol Evol.* - 2013b. - Vol. 5. - P. 2020-2031.
197. Zingler N., Weichenrieder O., Schumann G.G. APE-type non-LTR retrotransposons: determinants involved in target site recognition // *Cytogenet Genome Res.* - 2005. - Vol. 110. - P. 250-268.
198. Zhou L., Mitra R., Atkinson P.W., Hickman A.B., Dyda F., Craig N.L. Transposition of hAT elements links transposable elements and V(D)J recombination // *Nature.* - 2004. - Vol. 432. - P. 995-1001.
199. Zuccolo A., Sebastian A., Talag J., Yu Y., Kim H., Collura K., Kudrna D., Wing R.A. Transposable element distribution, abundance and role in genome size variation in the genus *Oryza* // *BMC Evol. Biol.* - 2007. - Vol. 7. - P. 152.
200. Zupunski V., Gubensek F., Kordis D. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Mol. Biol. Evol.* - 2001. - Vol. 18. - P. 1849–1863.

ПРИЛОЖЕНИЯ

Приложение 1

Таблица. Филогенетические группы мобильных генетических элементов, соответствующие праймеры и программы ПЦР амплификации, использованные в исследовании.

Филогенетическая группа МГЭ	Последовательности праймеров* (литературный источник)
Non-LTR ретротранспозоны	
Клад <i>CR1</i>	CR1-S: 5' -TATCTTCTTCTCCNGGNCCNGAYGG-3' CR1-A: 5'-CAAAAACACTGCCYTGNGGNACNCC-3' (Novikova <i>et al.</i> , 2007)
5'-UTR элемента <i>CR1B</i> из <i>M. teiuis</i>	MteCR1B-5'UTR-S: 5'- GTAGTAGTAGATCCAAGTCTGCAGTTCG-3' MteCR1B-ORF-A: 5'-TTATTGGGCTCAGTTTGCACCC-3.
<i>mariner</i> - like ДНК транспозоны	
Группа <i>mariner</i> - like элементов	MAR-124F: 5'-TGGGTNCCNCAYGARYT-3' MAR-276R: 5'-GGNGCNNARRTCNGG-3' (Robertson, 1993)
Элемент <i>Bmmar1</i> (семейство <i>mori</i>)	Bmmar1-ITR: 5'-CTTAGTCTGGCCATAAATACTGTTACAAAA-3' (Robertson, Asplund, 1996)
Элемент <i>Bmmar6</i> (семейство <i>mori</i>)	Bmmar6-ITR: 5'-MCTAGTCAGGTCATAARTWYTGTCAC-3' (Robertson, Walden, 2003)
Элемент <i>BmmarY</i> (подсемейство <i>vertumana</i>)	BmmarY-ITR: 5'-GGCTGCACTAAAAGTATCGGGA-3' (настоящее исследование)
Элемент <i>Bmmar3</i> (подсемейство <i>secropia</i>)	Bmmar3-ITR: 5'-CCTTACATATGAAATTAGCG-3' (Kumaresan, Mathavan, 2004)

* - Условные обозначения: Y = C+ T, M=A+ C, R=A+ G, W=A+T и N=A+G+ C+ T.

Таблица (продолжение). Филогенетические группы мобильных генетических элементов, соответствующие праймеры и программы ПЦР амплификации, использованные в исследовании.

Филогенетическая группа МГЭ	Последовательности праймеров* (литературный источник)
Химерные последовательности <i>CR1B/MLE</i>	
<i>CR1B/BmmarY-like</i>	BmCR1B-A1: 5'-TCTTCCACATCCGGCACACG-3' MacmarY-S1: 5'-CTTGTCAAACTTGTTTAGTCGTTGA-3' (настоящее исследование)
<i>CR1B/Bmmar1-like</i>	BmCR1B-A1: 5'-TCTTCCACATCCGGCACACG-3' Bmmar1-S1: 5'-TGTGTTTTCTCATTTTGGCGCCA-3' (настоящее исследование)
<i>CR1B/Bmmar1-like</i>	BmCR1B-A1: 5'-TCTTCCACATCCGGCACACG-3' Bmmar1-ITR: 5'-CTTAGTCTGGCCATAAATACTGTTACAAAA-3'
<i>CR1B/Bmmar6-like</i>	BmCR1B-A1: 5'-TCTTCCACATCCGGCACACG-3' Bmmar6-ITR: 5'-MCTAGTCAGGTCATAARTWYTGTCAC-3'

* - Условные обозначения: Y = C+ T, M=A+ C, R=A+ G, W=A+T и N=A+G+ C+ T.

Приложение 2

Таблица. Распространение различных подсемейств и семейств *mariner*-like TIR ДНК транспозонов в геномах представителей отряда Lepidoptera и класса Insecta.

Семейство элементов	Подсемейство	Распространение в геномах Insecta (GenBank №)	Распространение в геномах Lepidoptera (GenBank №)
<i>mariner</i>	<i>DTTMarCRI</i>	<i>B. mori</i> (AF526192; AF526191)	<i>Andrena erigenia</i> (U91348); <i>Apis mellifera</i> (XM001122397); <i>Apicauta funebris</i> (U91358)
	<i>cecropia</i>	<i>B. mori</i> (<i>Bmmar3</i> , <i>Bmmar5</i>); <i>B. mandarina</i> (AB237579, AB473777, AB473782, AB23754); <i>Papilio xuthus</i> (AB055185); <i>Cephonodes hylas</i> (AB055183); <i>Attacus atlas</i> (AB006464); <i>Hyphantria cunea</i> (AB055186); <i>Junonia almana</i> (AB055182); <i>Antheraea yamamai</i> (AB041902); <i>Danaus chrysippus</i> (AB055181); <i>Hyalophora cecropia</i> (L10444); <i>Saturnia japonica</i> (AB085937); <i>Ephestia cautella</i> (L10484); <i>Depressaria pastinacella</i> (U91380)	<i>Orthotrichia cristata</i> (U91373)
	<i>mellifera</i>	<i>B. mori</i> (<i>Bmmar4</i>); <i>Bombyx mandarina</i> (AF212134); <i>Ephestia cautella</i> (L10481); <i>Atteva punctella</i> (U91342); <i>Samia cynthia</i> (AB041897)	<i>Chrysopus vittatus</i> (L10502); <i>Orthotrichia cristata</i> (U913375)

Таблица (продолжение). Распространение различных подсемейств и семейств *mariner*-like TIR ДНК транспозонов в геномах представителей отряда Lepidoptera и класса Insecta.

Семейство элементов	Подсемейство	Распространение в геномах Insecta (GenBank №)	Распространение в геномах Lepidoptera (GenBank №)
<i>mariner</i>	<i>vertumana</i>	<i>Heliothis virescens</i> (DQ174779)	<i>Bactrocera neohumeralis</i> (AF348438); <i>Teleopsis rubicunda</i> (EF035477); <i>Teleopsis quinqueguttata</i> (DQ197025)
	<i>irritans</i>	<i>Adoxophyes honmai</i> (AB020617)	<i>Haematobia irritans</i> (U11642, U11641); <i>Chrysoperla plorabunda</i> (U11654)
	<i>mauritiana</i>	<i>Mamestra brassicae</i> (AF465247)	<i>Drosophila mauritiana</i> (X78906); <i>Drosophila simulans</i> (X78907); <i>Cerotoma trifurcata</i> (L10460); <i>Oncopeltus fasciatus</i> (L10489); <i>Bobmus terrestris</i> (AJ312716)
	<i>rosa</i>	--	<i>Ceratitis rosa</i> (AY034623, AY034620)
	<i>capitata</i>	--	<i>Diasemopsis dubia</i> (EF407577); <i>Buenoa sp.</i> (U91351)
	<i>lineata</i>	--	<i>Ctenolepisma lineate</i> (L10493); <i>Nabis sp.</i> (U91362)
	<i>DTTMarUrt</i>	--	<i>Ctenolepisma lineata</i> (L10495)

Таблица (продолжение). Распространение различных подсемейств и семейств *mariner*-like TIR ДНК транспозонов в геномах представителей отряда Lepidoptera и класса Insecta.

Семейство элементов	Подсемейство	Распространение в геномах Insecta (GenBank №)	Распространение в геномах Lepidoptera (GenBank №)
<i>mori</i>	<i>mori</i>	<i>B. mori</i> (<i>Bmmar1</i> , <i>Bmmar6</i>), <i>B. mandarina</i> (AY172028, FJ897530, AB363010, AB363017, AB363018)	--
<i>ludens</i>	<i>ludens</i>	--	<i>Anastrepha ludens</i> (AY034627, AY034626)