

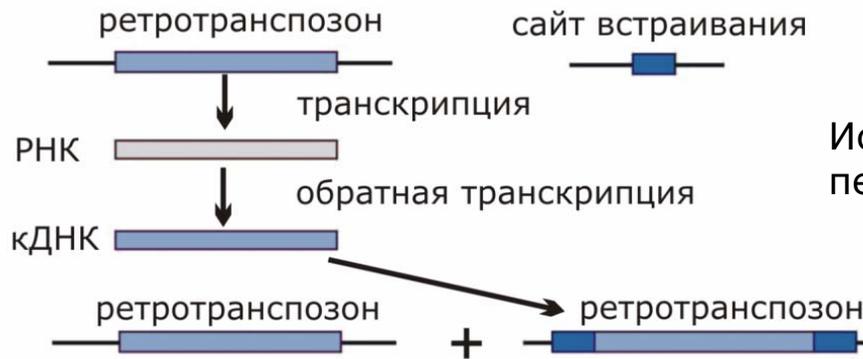
Использование экспериментального и биоинформатического подходов для исследования ретротранспозонов в геномах эукариот

Ольга Новикова

Новосибирский государственный университет
Институт цитологии и гентики СО РАН

Новосибирск 2007

Ретротранспозоны

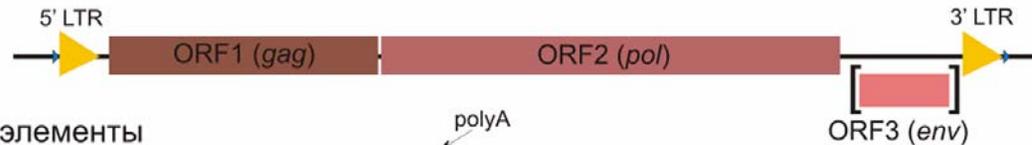


Используют в процессе своего перемещения **обратную транскрипцию**.

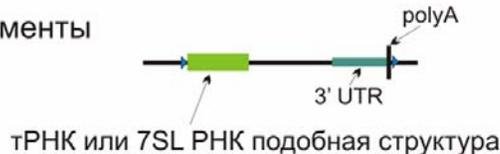
Non-LTR ретротранспозоны



LTR ретротранспозоны



SINE элементы



▶ - TSD

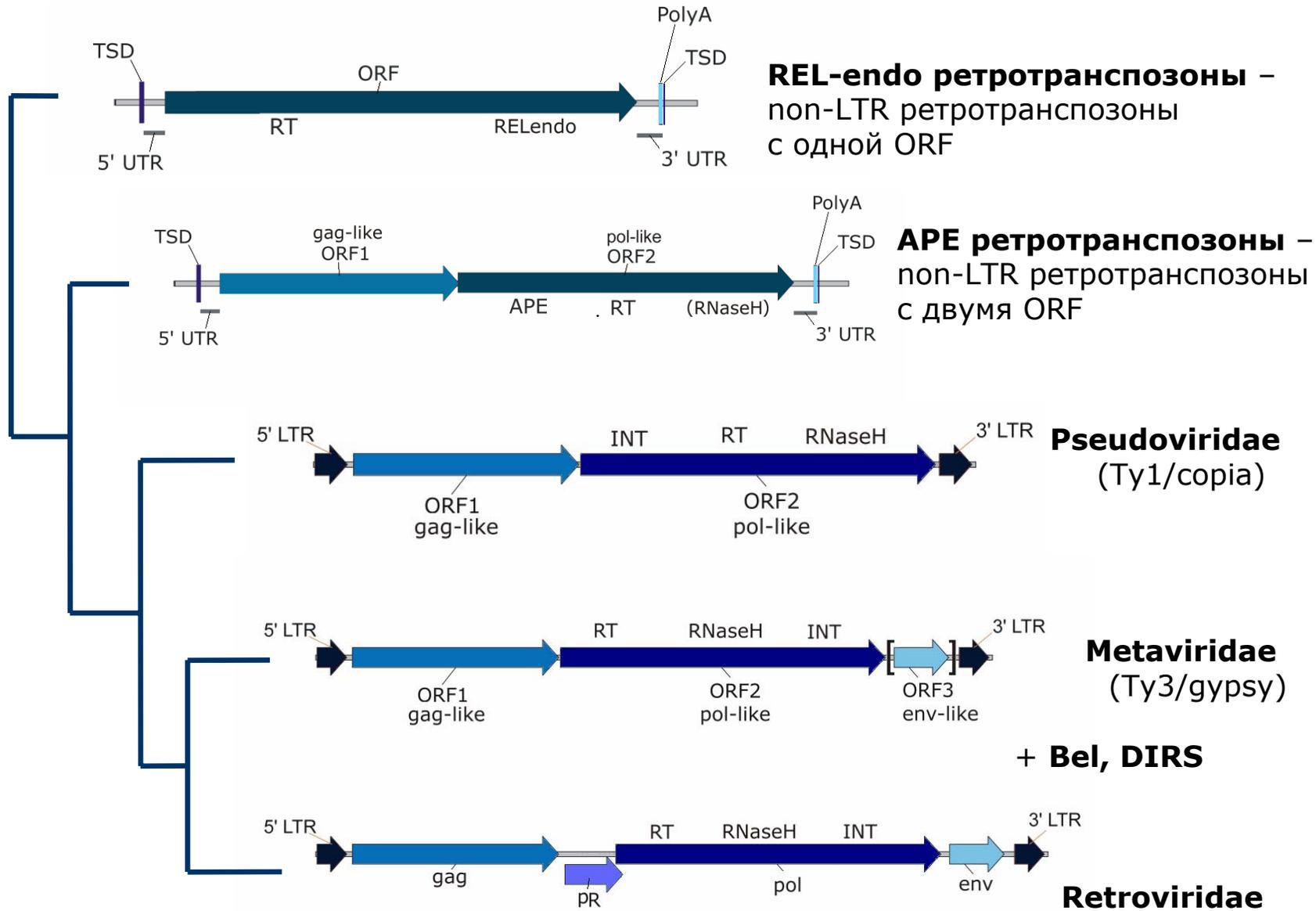
ORF - открытая рамка считывания

5'UTR и **3'UTR** - 5' и 3' нетранслируемые районы

5'LTR и **3'LTR** - 5' и 3' длинные концевые повторы

TSD - дублированные повторы-мишени

Эволюция ретротранспозонов



Ретротранспозоны: эволюция и распределение

Цель: изучение распространения и эволюции ретротранспозонов в геномах эукариотических организмов.

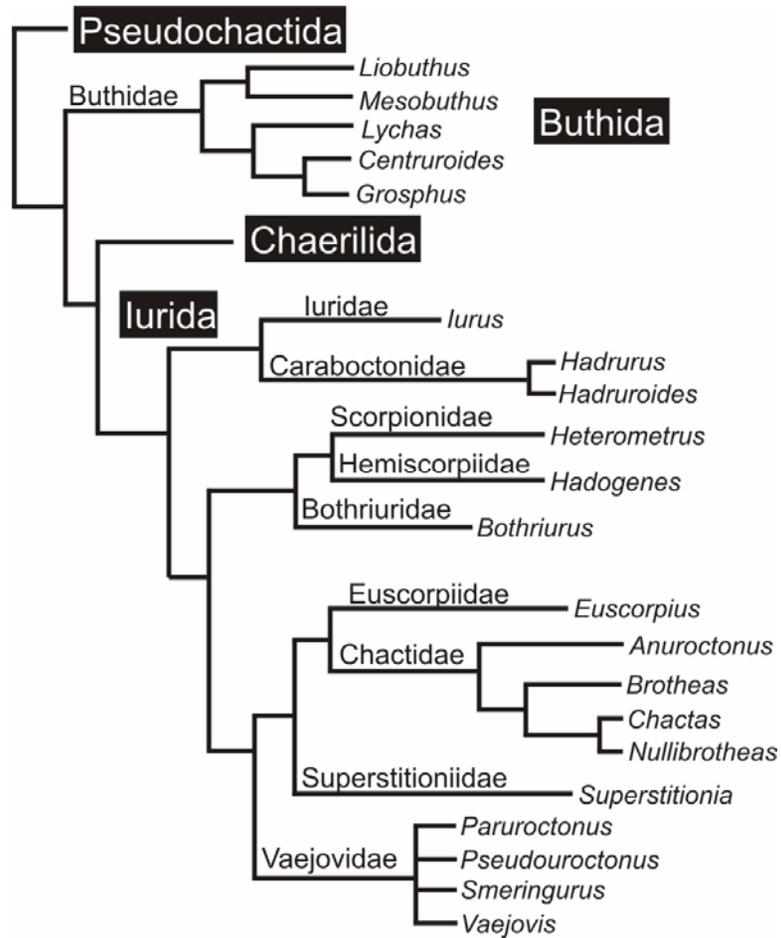
Задачи:

1. Исследование разнообразия non-LTR ретротранспозонов среди представителей отряда Scorpiones (Arachnida: Scorpiones).
2. Исследование разнообразия non-LTR ретротранспозонов среди представителей рода *Maculinea* (Lepidoptera: Lycaenidae).
3. Определение нуклеотидных последовательностей элементов, относящихся к различным группам non-LTR ретротранспозонов, из видов отряда Scorpiones и рода *Maculinea*.
4. Обработка последовательностей, изучение внутри- и межвидовой вариабельности, оценка разнообразия групп non-LTR ретротранспозонов в геномах видов отряда Scorpiones и рода *Maculinea*. Построение филогенетических деревьев non-LTR ретротранспозонов.

Ретротранспозоны: эволюция и распределение

5. Исследование разнообразия LTR ретротранспозонов с помощью биоинформатических подходов в геномах восьми эукариотических организмов, включая три вида грибов (Fungi) и пять видов животных (Animalia).
6. Исследование разнообразия и получение полных последовательностей LTR ретротранспозонов из геномов *Aspergillus fumigatus* и *A. nidulans* (Fungi: Ascomycota). Анализ нуклеотидных последовательностей, построение филогенетических деревьев LTR ретротранспозонов.
7. Исследование разнообразия и получение полных последовательностей LTR ретротранспозонов из генома *Phanerochaete chrysosporium* (Fungi: Basidiomycota). Анализ нуклеотидных последовательностей, построение филогенетических деревьев LTR ретротранспозонов.

Non-LTR ретротранспозоны Scorpiones



2 подотряда
10 семейств
21 род
22 вида скорпионов

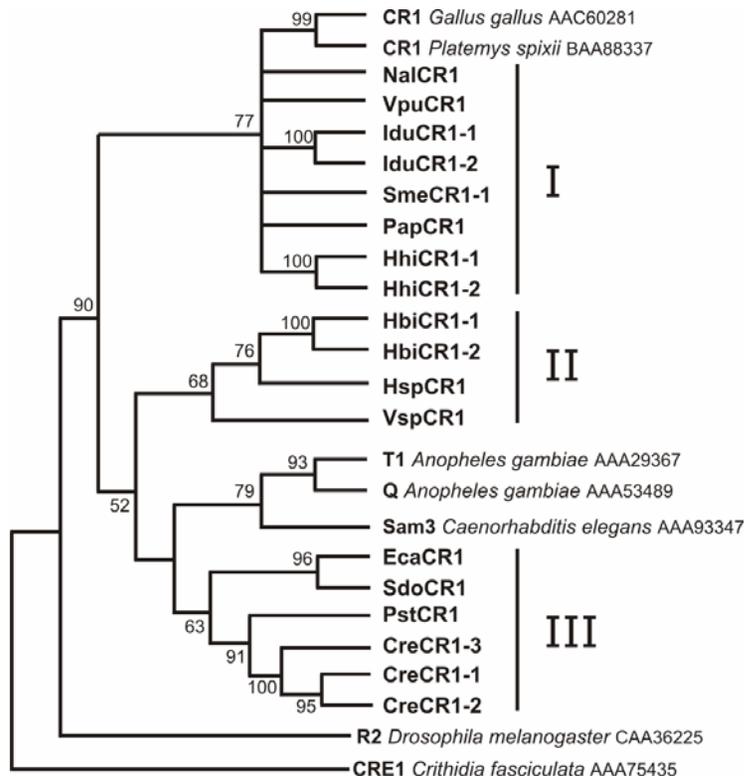
- Проведен ПЦР-анализ с использованием вырожденных праймеров, клонирование и секвенирование полученных продуктов.
- Полученные последовательности были использованы для сравнительного и филогенетического анализов

Материал предоставлен Проф. Виктором Фетом (Marshall University, USA)

Три филогенетические группы были обнаружены в геномах Scorpiones **CR1, Jockey** и **I**

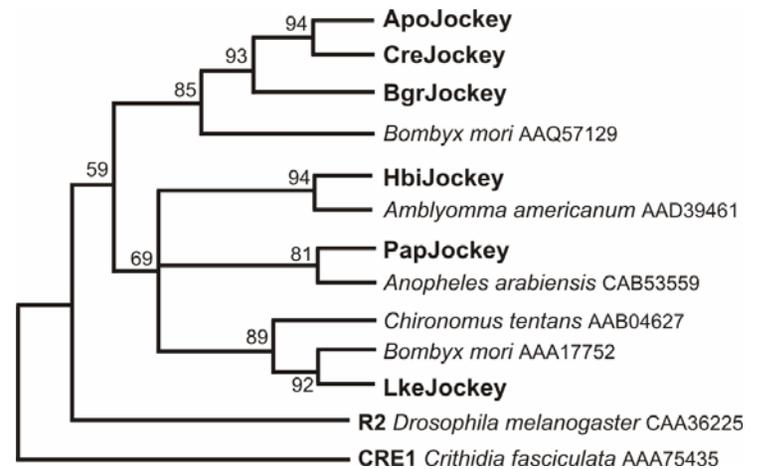
Non-LTR ретротранспозоны Scorpiones

CR1 филогенетическая группа

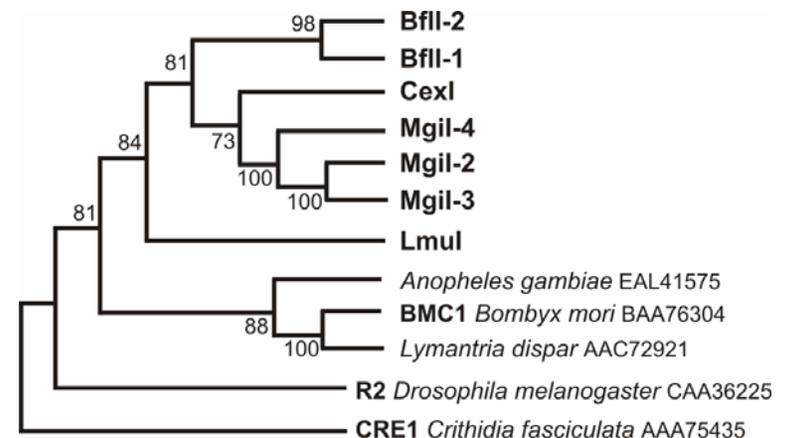


Подавляющее большинство фрагментов элементов представляли собой последовательности нарушенной RT.

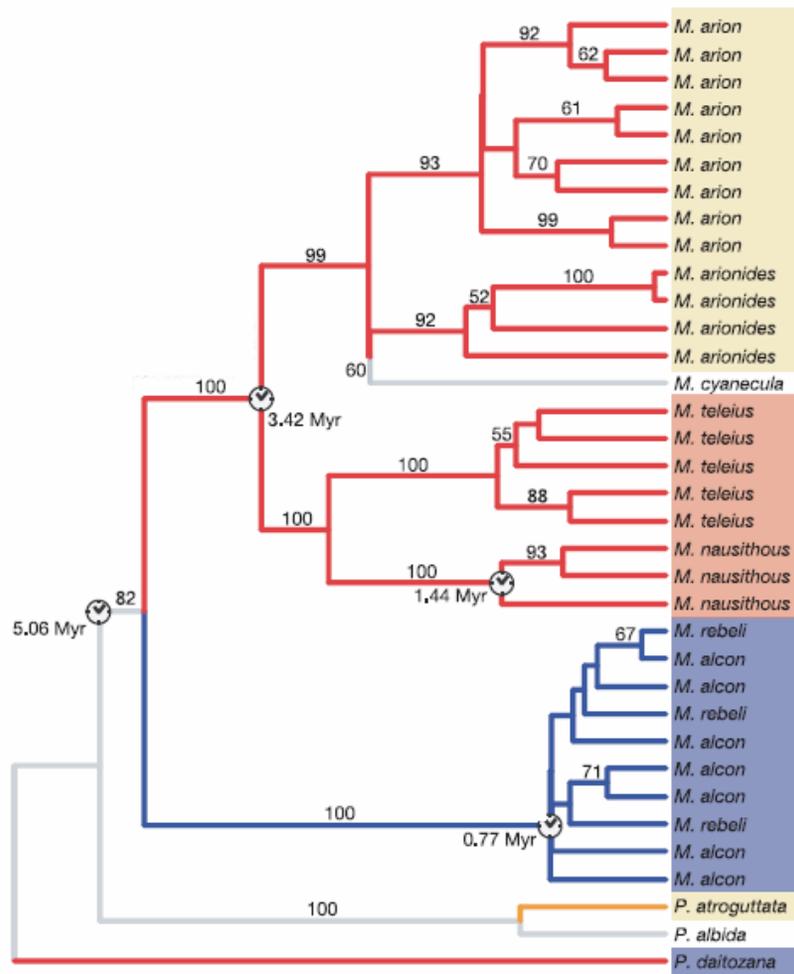
Jockey филогенетическая группа



I филогенетическая группа



Non-LTR ретротранспозоны *Maculinea*



Представители рода *Maculinea* являются мирмекофилами и являются одним из интереснейших объектов изучения для экологов. Относятся к исчезающим видам.

4 вида:

Maculinea teleius
Maculinea nausithous
Maculineaalcon
Maculinea arion

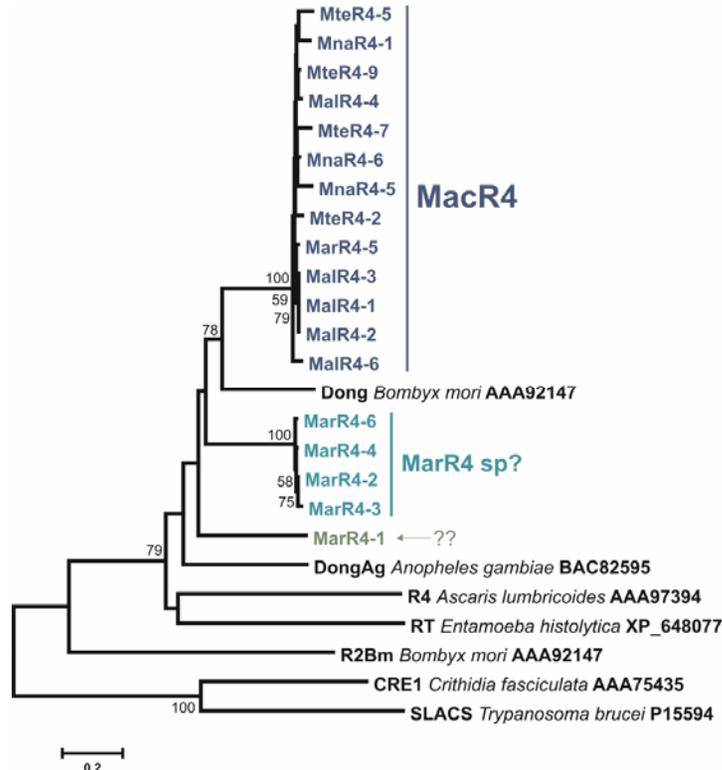
Материал предоставлен Проф. Михалом Войциховски (Jagiellonian University, Poland).

- Проведен ПЦР-анализ с использованием вырожденных праймеров, клонирование и секвенирование полученных продуктов.
- Полученные последовательности были использованы для сравнительного и филогенетического анализов

Четыре филогенетические группы были обнаружены в геномах *Maculinea* R4, Jockey, CR1 и R1

Non-LTR ретротранспозоны *Maculinea*

R4 филогенетическая группа

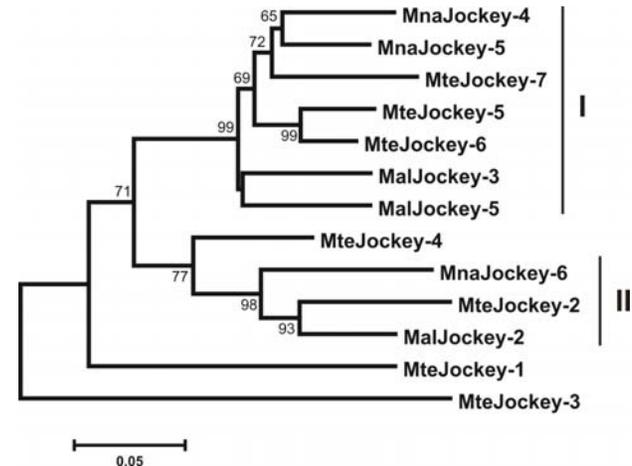


R4-подобные элементы обнаружены у всех исследованных *Maculinea* - семейство MacR4.

Элементы MacR4 семейства имеют высокое сходство как внутри видов, так и между видами.

Два семейства специфичны для *M. arion*.

Jockey филогенетическая группа

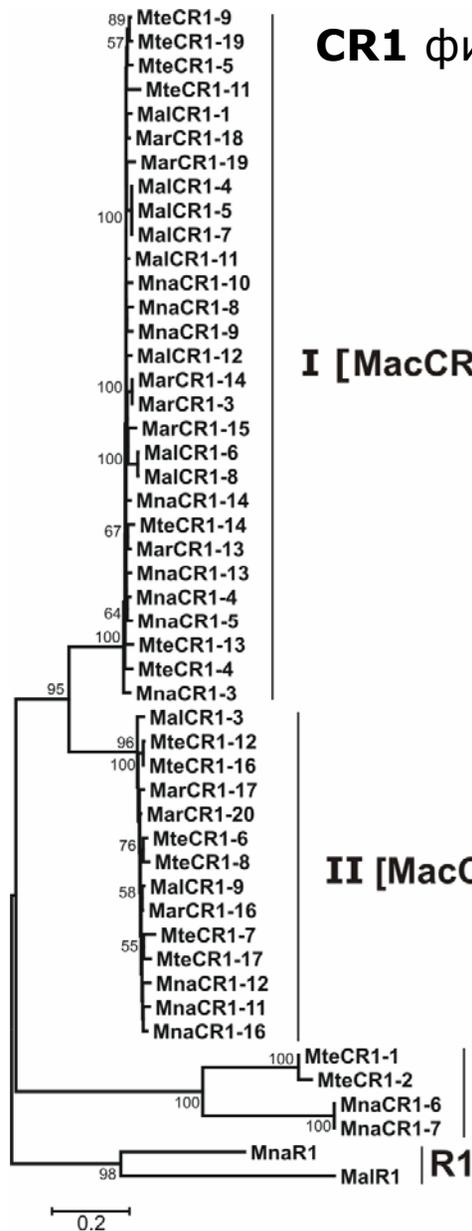


Jockey-подобные элементы обнаружены у всех исследованных *Maculinea*, за исключением *M. arion*.

Jockey-подобные элементы разнородны, сходство на нуклеотидном уровне составляло в среднем 74%.

Подавляющее большинство фрагментов R4 и Jockey элементов представляли собой последовательности интактной RT.

Non-LTR ретротранспозоны *Maculinea*



CR1 филогенетическая группа

Три семейства в составе CR1

MacCR1A

MacCR1B

MacCR1C (T1Q)

I [MacCR1A]

Сходство *MacCR1A* и *MacCR1B*

элементов составляло в среднем **73.5%** нуклеотидной последовательности и **88%** на аминокислотном уровне.

II [MacCR1B]

BLAST анализ показал наличие *MacCR1B*-подобных ретроэлементов в геноме *Bombyx mori*, причем сходство между *MacCR1B* и *BmCR1B* элементами составляло **более 96%** на нуклеотидном уровне.

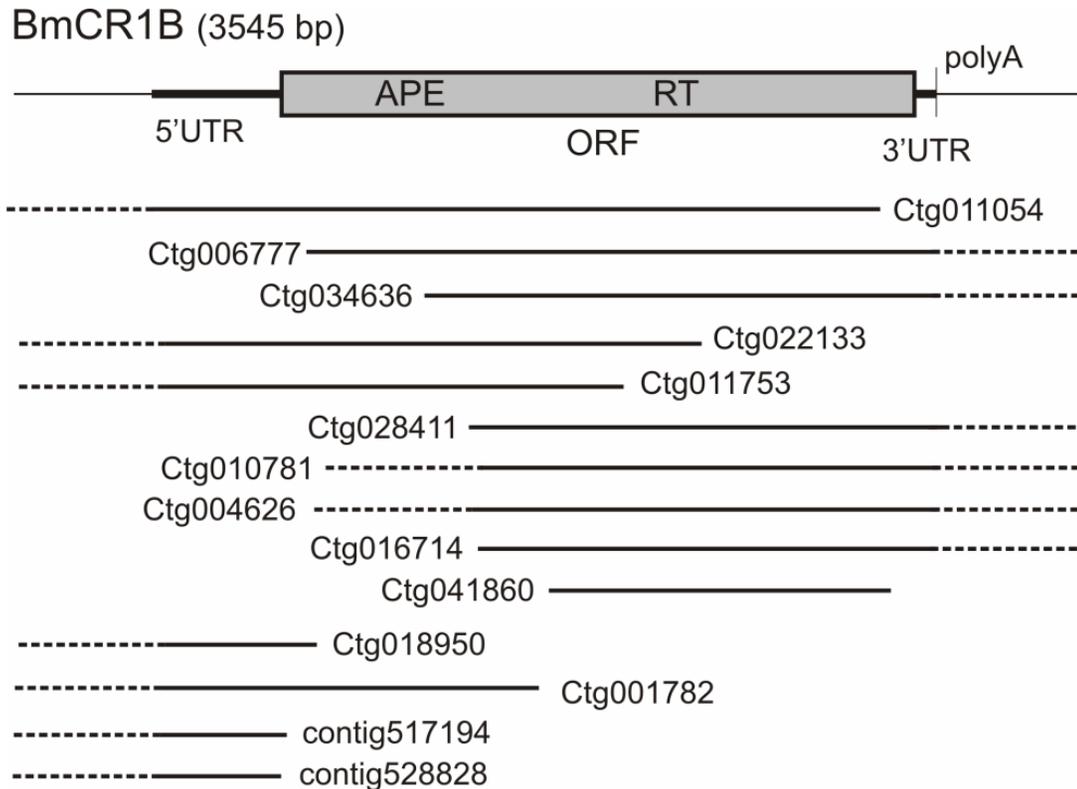
III [MacCR1C = T1Q]

R1 филогенетическая группа

0.2

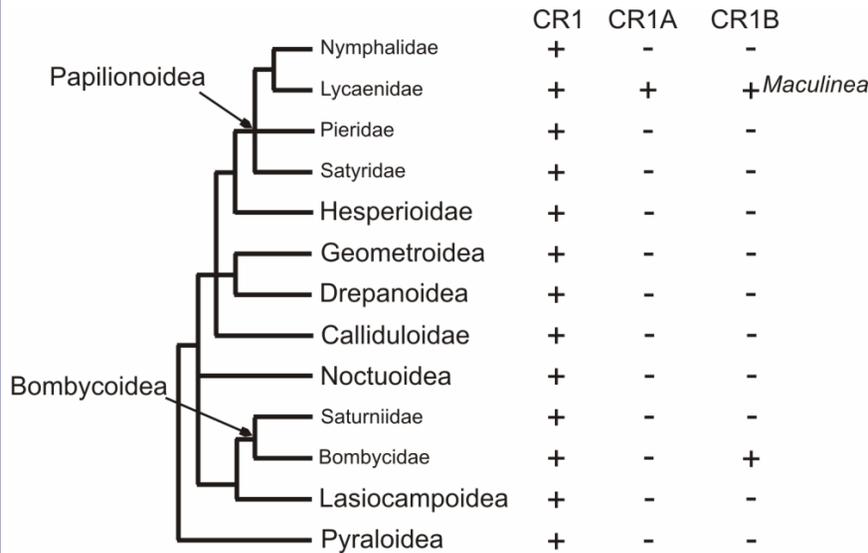
BmCR1B non-LTR ретротранспозон *Bombyx mori*

BmCR1B ретроэлемент был реконструирован на основе выравнивания геномных фрагментов *B. mori*.



ORF - открытая рамка считывания
5'UTR и **3'UTR** - 5' и 3' нетранслируемые районы
RT - домен обратной транскриптазы
APE - домен апурин/апиримидин эндонуклеазы

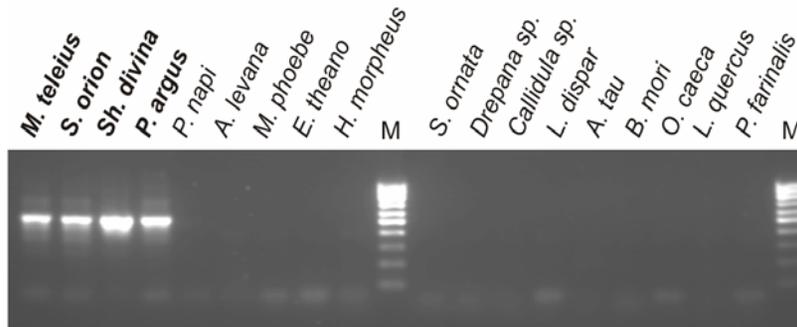
CR1A и CR1B семейства в геномах Lepidoptera



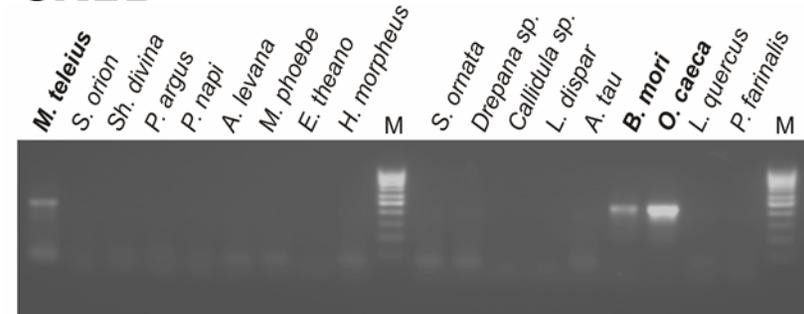
Для исследования рапределения семейств CR1A и CR1B:

- Были выбраны 17 представителей отряда Lepidoptera, подотряда Ditrysia, относящиеся к 9-ти семействам.
- Были выбраны специфичные праймеры для семейств элементов CR1A и CR1B.
- Проведен ПЦР-анализ, клонирование и секвенирование полученных продуктов.

CR1A



CR1B



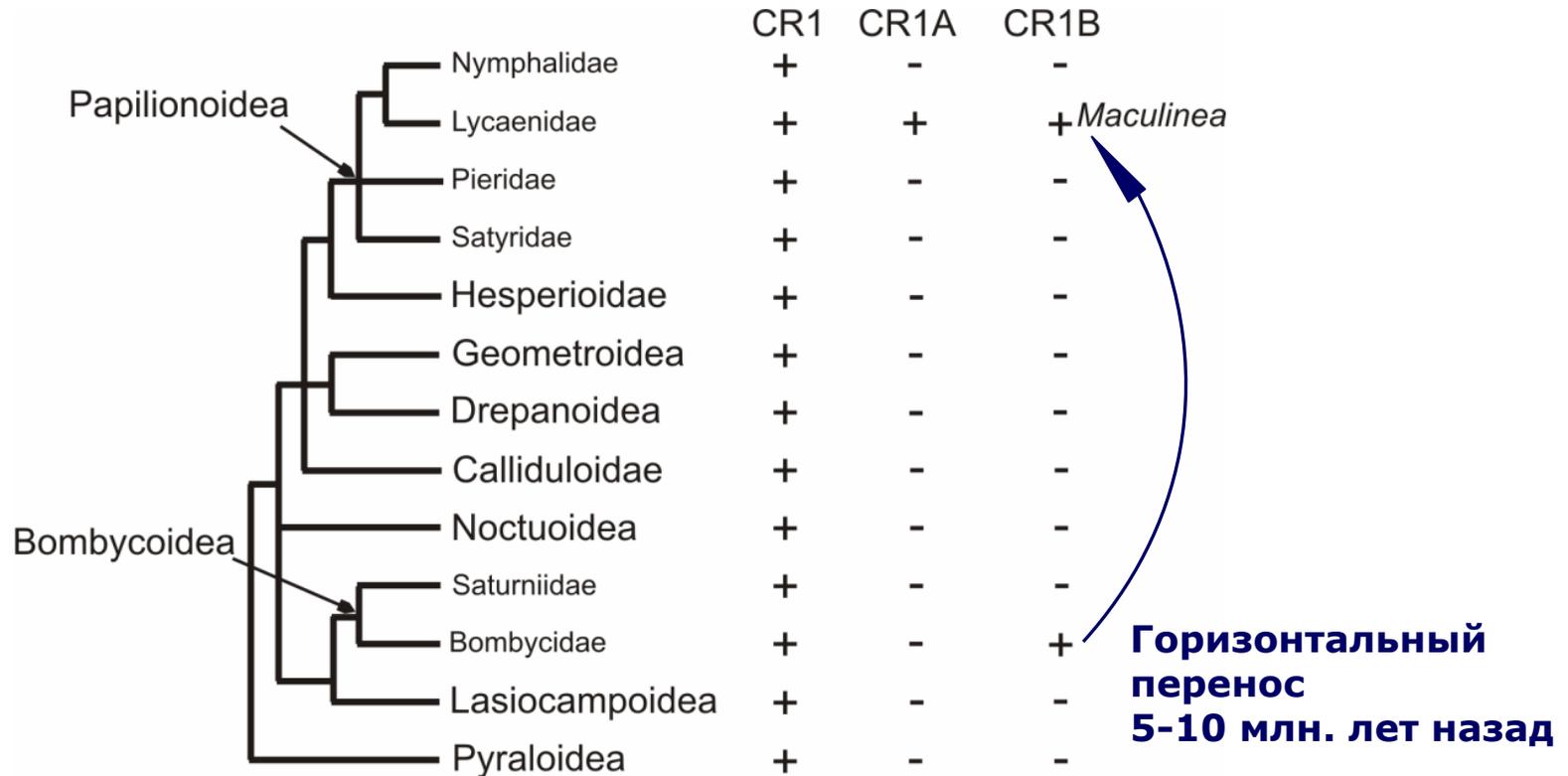
Материал предоставлен к.б.н. Олегом Костериным (ИЦиГ) и к.б.н. Владимиром Дубатовым (ИСиЭЖ).

Горизонтальный перенос *CR1B* элементов

Горизонтальный перенос – это процесс, благодаря которому может осуществляться перемещение генов у репродуктивно изолированных видов. Механизм неизвестен.

Критерии:

- Неравномерное распределение - присутствие у эволюционно далеких видов.
- Скорость накопления замен в последовательностях мобильных элементов ниже, чем в последовательностях функциональных генов.



LTR ретротранспозоны в геномах эукариот



UniPro GenomeBrowser - <http://genome.unipro.ru/>
Plugins - <http://genome.sourceforge.net/>



Алгоритм HMM (Hidden Markov Model), основанный на скрытых цепях Маркова¹
Приложение **HMM Search**.

Поиск мобильных элементов по заданным характеристикам – наличие прямых или обратных повторов, наличие открытых рамок считывания.
Приложение **Find ME**.

UniPro GenomeBrowser

MEGA 3.0

он-лайн сервисы NCBI

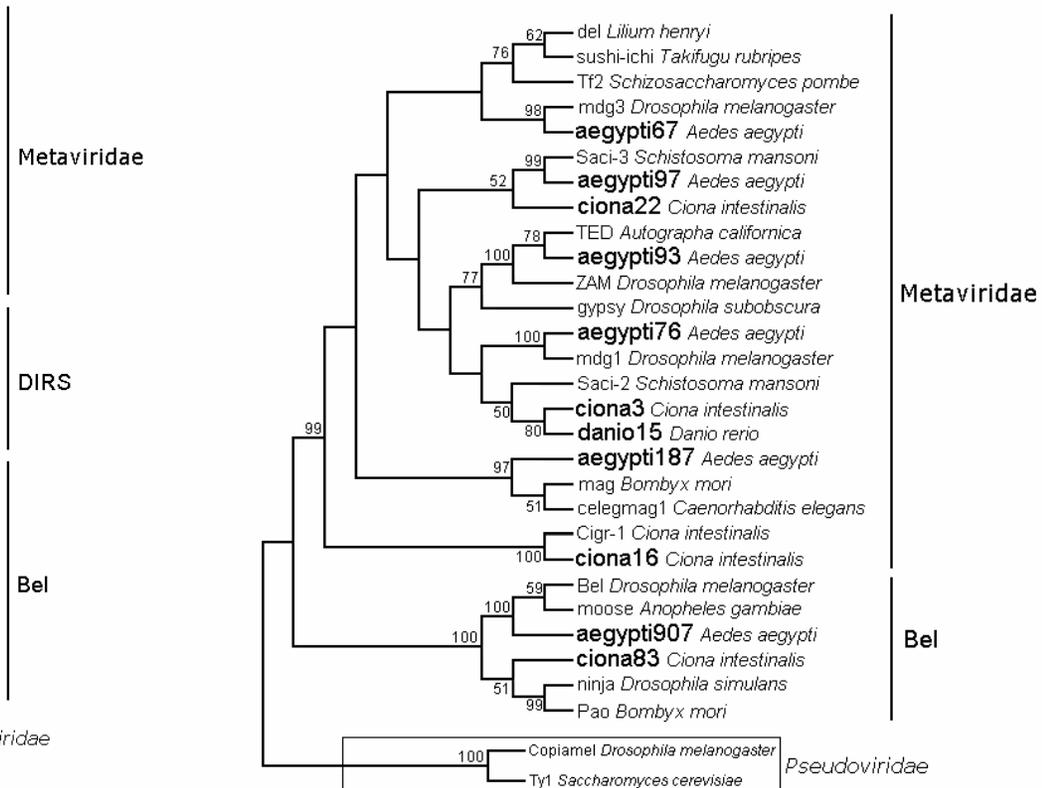
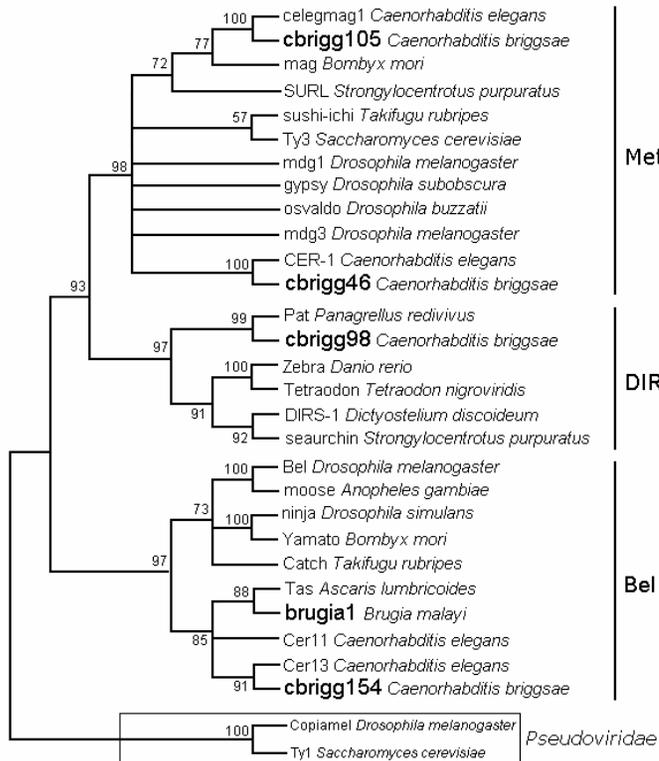
LTR ретротранспозоны в геномах эукариот



- ▶ *Caenorhabditis briggsae* - **124 HMM сигнала**
- Brugia malayi* - **11 HMM сигналов**
- Aedes aegypti* - **166 HMM сигналов**
- Ciona intestinalis* - **346 HMM сигналов**
- Danio rerio* - **210 HMM сигналов**

▶ **15 новых** LTR элементов, относящихся к группам **Metaviridae, DIRS** и **Bel**.

- UniPro GenomeBrowser
- MEGA 3.0
- он-лайн сервисы NCBI



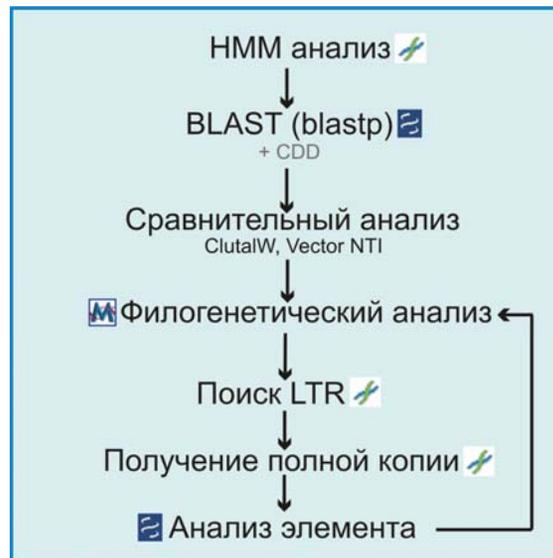
LTR ретротранспозоны *Aspergillus fumigatus* и *Aspergillus nidulans*



▶ *Aspergillus fumigatus* - основной возбудитель инвазивных микозов. Более 90% случаев.



▶ *Aspergillus nidulans* - один из важнейших модельных организмов для исследований биологии клетки и регуляции генов.



▶ *A. fumigatus* - **60 HMM сигналов**
A. nidulans - **25 HMM сигналов**

▶ Все элементы - **Metaviridae**
A. fumigatus - **4 группы** элементов
A. nidulans - **2 группы** элементов

▶ *A. fumigatus* и *A. nidulans* - большинство копий элементов имеют нарушения и замены C:G на T:A

UniPro GenomeBrowser

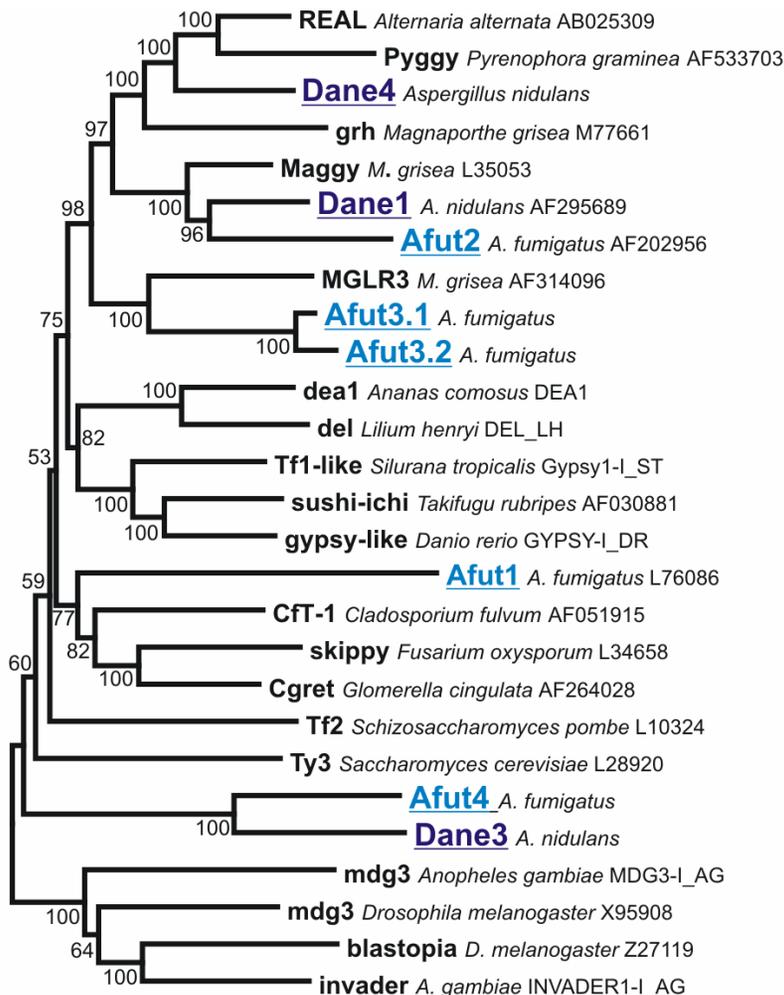
MEGA 3.0

он-лайн сервисы NCBI

LTR ретротранспозоны *Aspergillus fumigatus* и *Aspergillus nidulans*

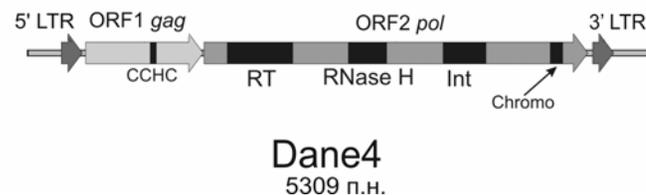
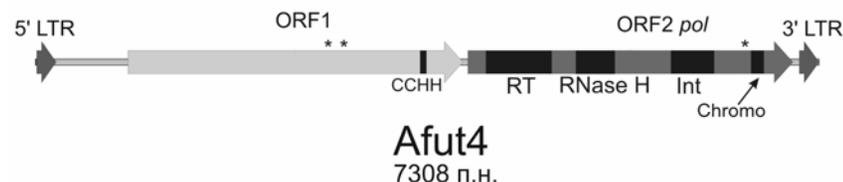
Aspergillus fumigatus:

Afut1¹, Afut2², **Afut3**, **Afut4**



Aspergillus nidulans:

Dane3, **Dane4** + Dane1, Dane2³



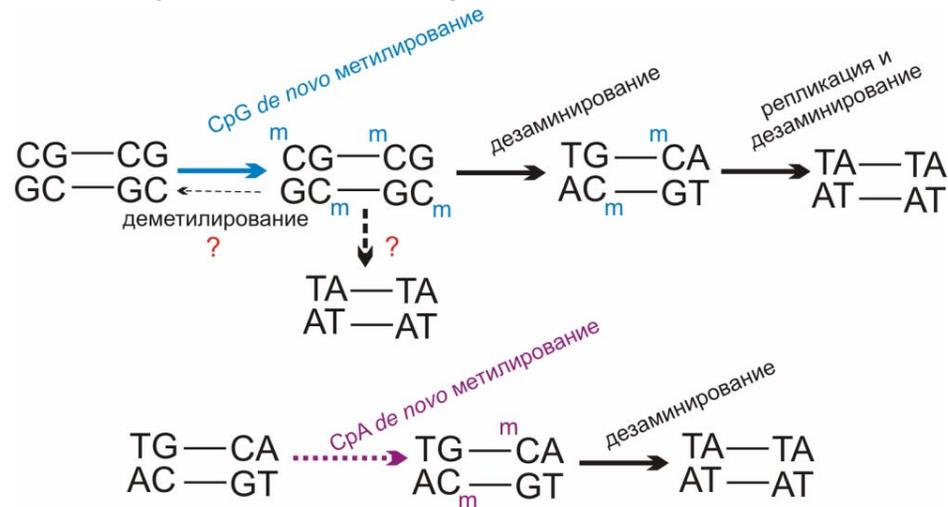
ORF - открытая рамка считывания; **5'LTR** и **3'LTR** - 5' и 3' длинные концевые повторы; **RT** - домен обратной транскриптазы; **RNase H** - домен рибонуклеазы H; **Int** - интегразы; **CCHH** и **CCHC** - цинковые пальцы; **Chromo** - хромодомен; * - стоп-кодоны.

1. Neugeglise et al. 1996.
2. Paris S, Latge JP. 2001.
3. Nielsen et al. 2001.

Гомолог-зависимая инактивация повторенных последовательностей грибов

Neurospora crassa - **RIP** инактивация.¹

Мутагенез, индуцированный повторяющимися последовательностями (RIP – repeat induced point mutations)



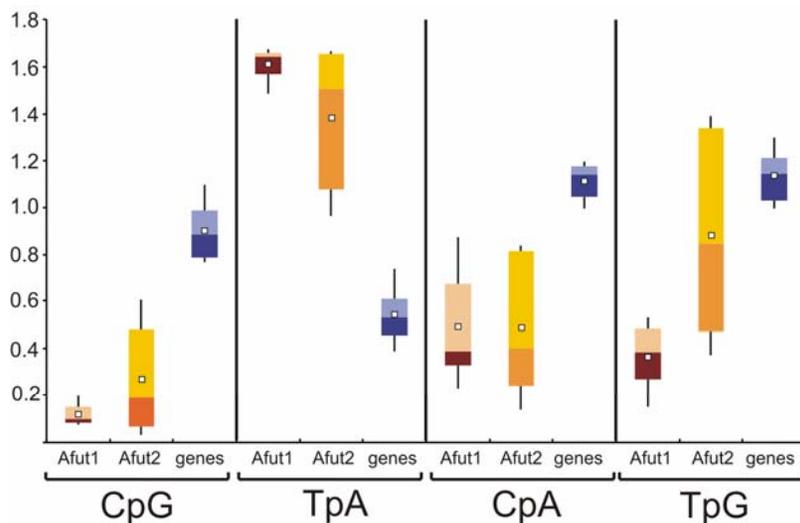
RIP-анализ

- Подсчет частот встречаемости CpG, CpA, TpG, TpA динуклеотидов.
- Сравнение частот встречаемости между генами и LTR ретротранспозонами.
- Реконструкция "de-RIP" элемента.

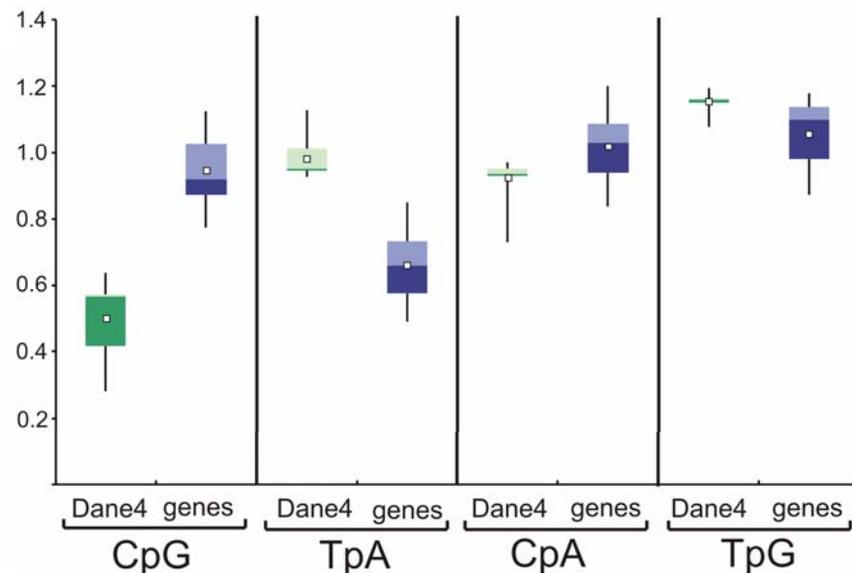
$F[\text{CpG}] = O[\text{CpG}] / E[\text{CpG}]$,
 где $O[\text{CpG}]$ – наблюдаемое число сайтов CpG в последовательности, $E[\text{CpG}]$ – ожидаемое число CpG сайтов,

$E[\text{CpG}] = [C] \times [G] / L$,
 где $[C]$ и $[G]$ – число C и G в последовательности, L – длина последовательности

Инактивация LTR ретротранспозонов *Aspergillus fumigatus* и *Aspergillus nidulans*



Бокс-плот частот встречаемости CpG, TpA, CpA и TpG сайтов в последовательностях *Afut1*, *Afut2* и генов из *Aspergillus fumigatus*



Бокс-плот частот встречаемости CpG, TpA, CpA и TpG сайтов в последовательностях *Dane4* и генов из *Aspergillus nidulans*

Сравнительный анализ частот встречаемости CpG, TpA, CpA и TpG сайтов для ретроэлементов и генов показал, что элементы *Afut1* и *Afut2* из *A. fumigatus* и *Dane4* из *A. nidulans* подверглись RIP инактивации по CpG и CpA сайтам.

LTR ретротранспозоны

Phanerochaete chrysosporium



▶ *Phanerochaete chrysosporium* - один из немногих грибов-базидиомицетов (Basidiomycota), способных разлагать лигнин, главный компонент клеточной стенки растений.



UniPro GenomeBrowser

MEGA 3.0

он-лайн сервисы NCBI

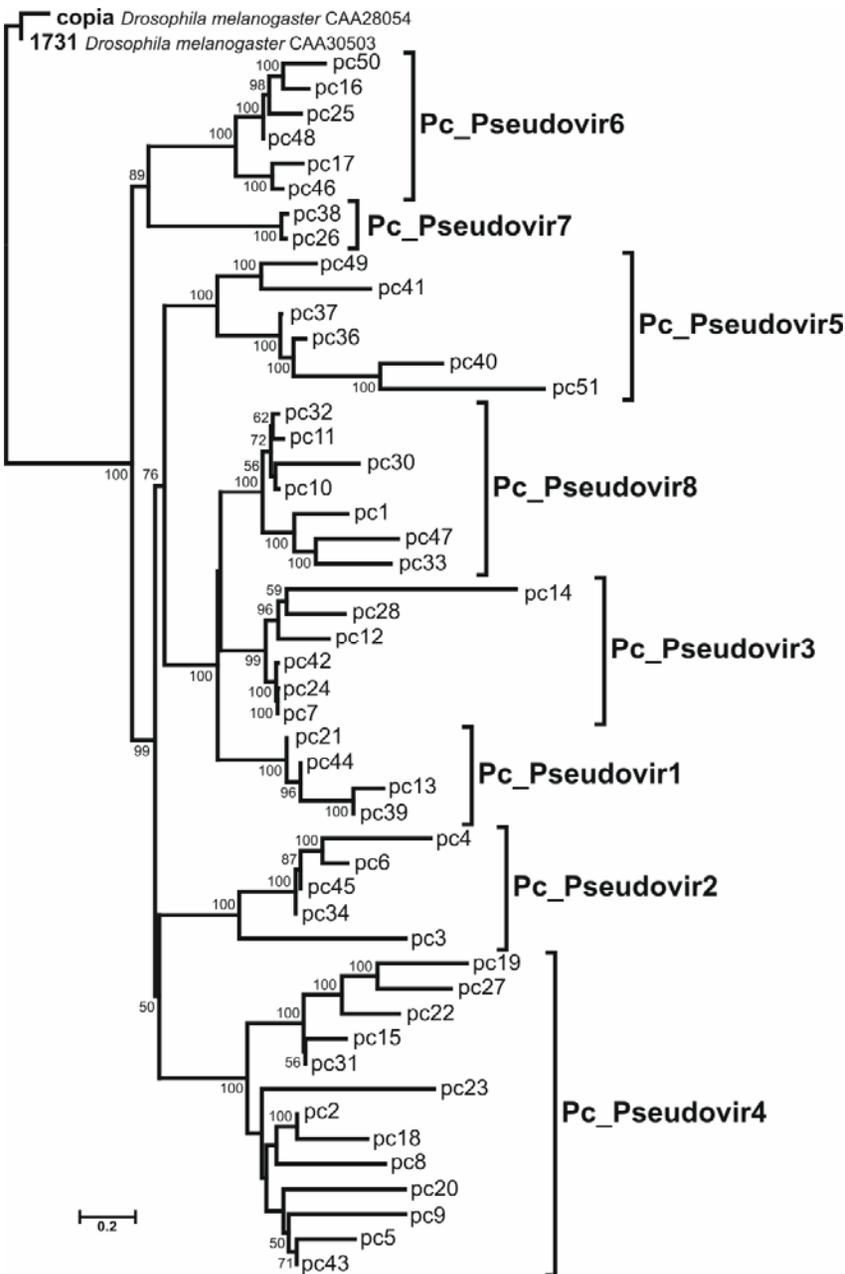
▶ Pseudoviridae - **51 HMM сигнал**
Metaviridae - **144 HMM сигналов**

▶ Pseudoviridae - **8 Pс_Pseudovir элементов**
Metaviridae - **15 Pс_Metavir элементов**

▶ Pseudoviridae - **6 элементов** представлены полноразмерными интактными копиями

Metaviridae - **7 элементов** представлены полноразмерными интактными копиями

Pseudoviridae из *Phanerochaete chrysosporium*



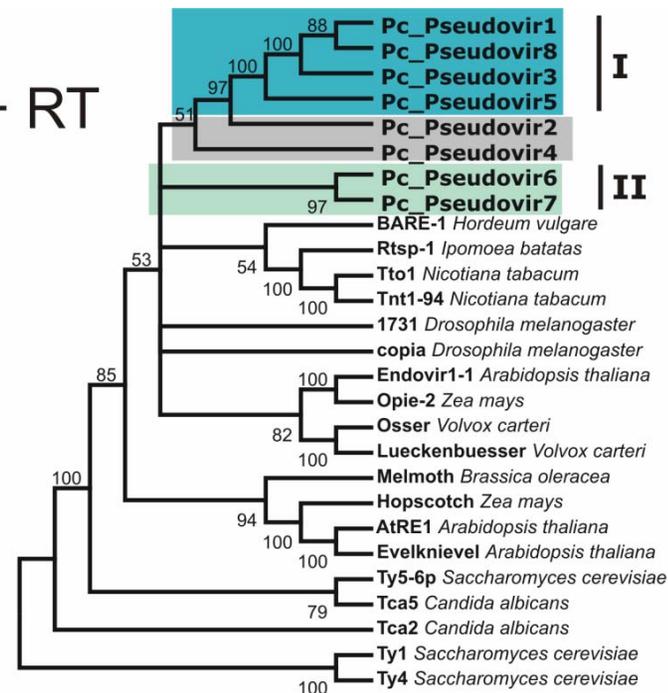
Копийность от 2 до 10 и более копий на геном.

Гомология между копиями элементов составляла от 85% до 99% нуклеотидной последовательности.

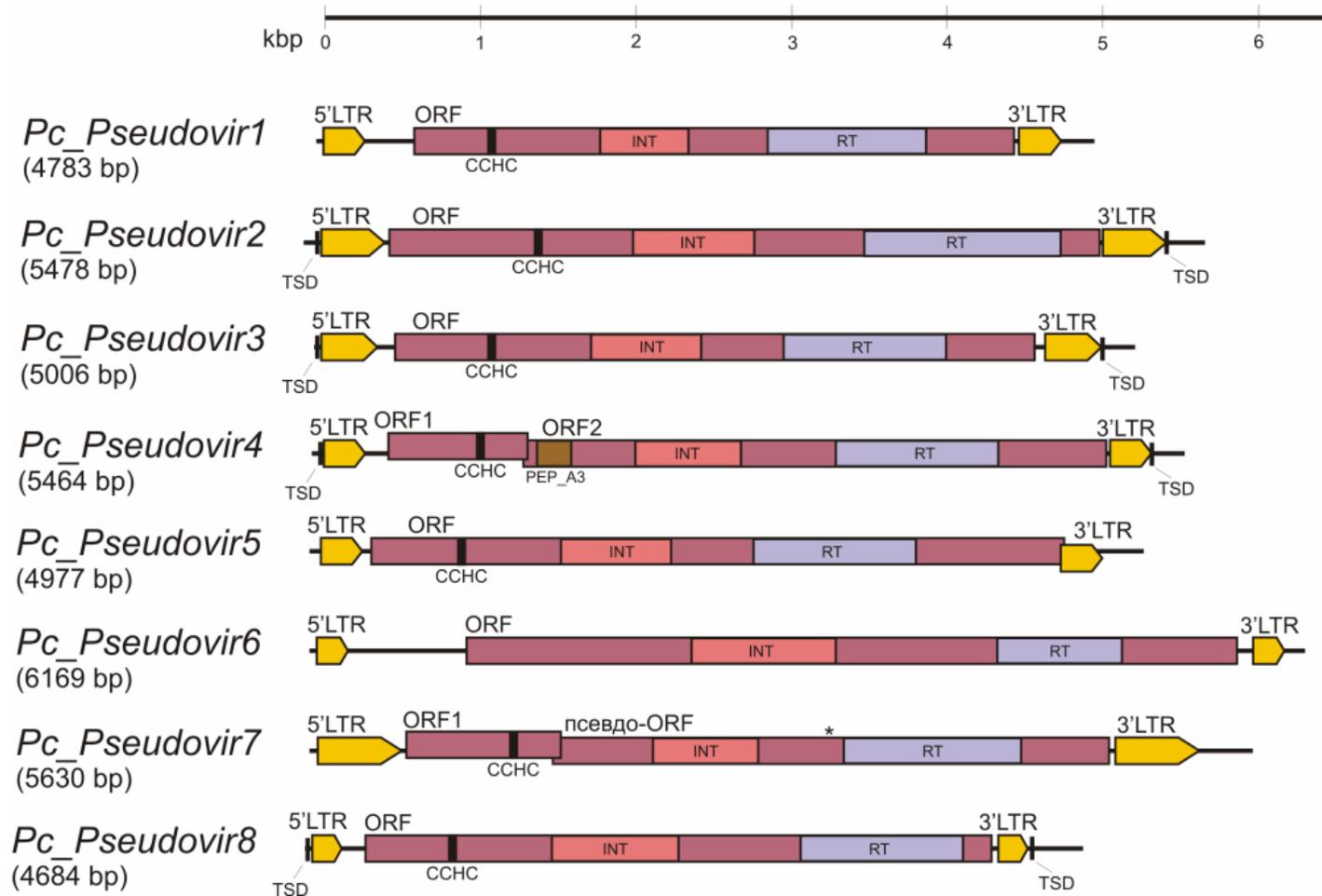
Гомология между 5' и 3' LTR составляла от 91.2% до 100%.

Обнаружены соло LTR (кроме *Pc_Pseudovir1* и *Pc_Pseudovir2*).

Int + RT

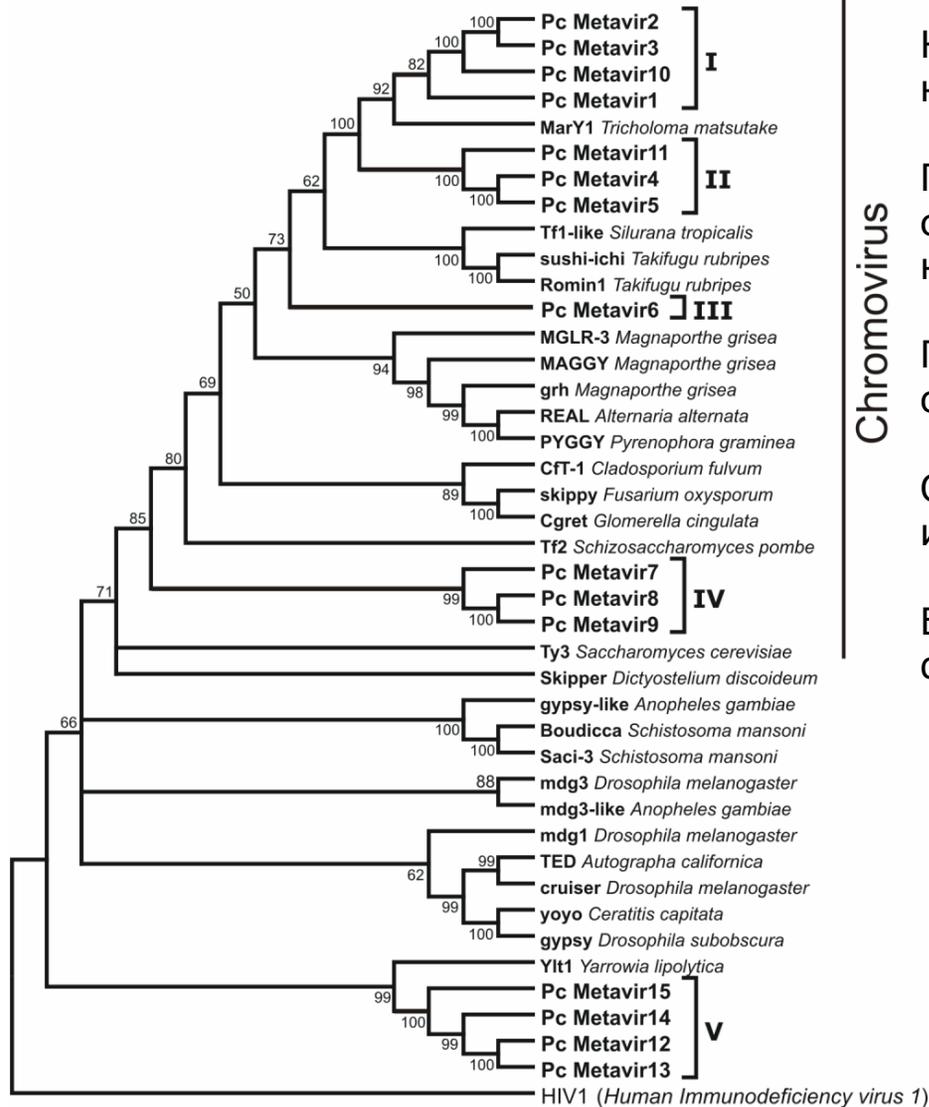


Pseudoviridae из *Phanerochaete chrysosporium*



ORF - открытая рамка считывания; **5'LTR** и **3'LTR** - 5' и 3' длинные концевые повторы; **RT** - домен обратной транскриптазы; **Int** - интегразы; **CCHC** - цистеиновый мотив; **PEP_A3** - протеаза; * - стоп-кодон; **TSD** - повторы-мишени.

Metaviridae из *Phanerochaete chrysosporium*



Копийность от 1 до 50 и более копий на геном.

Гомология между копиями элементов составляла от 82% до 99% нуклеотидной последовательности.

Гомология между 5' и 3' LTR составляла от 90.4% до 100%.

Обнаружены соло LTR *Pc_Metavir1,2,3* и 5.

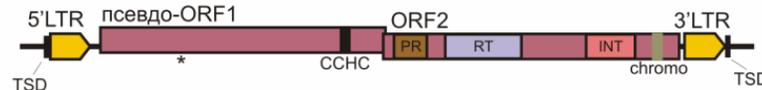
Большинство *Pc_Metavir* относятся к филогенетической группе Chromovirus.

Chromovirus

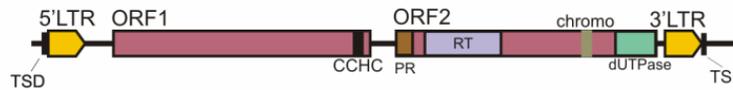
Metaviridae из *Phanerochaete chrysosporium*



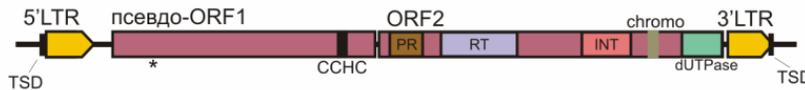
Pc_Metavir1
(7737 bp)



Pc_Metavir2
(7184 bp)



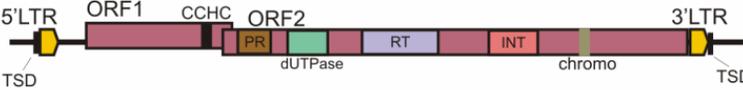
Pc_Metavir3
(8108 bp)



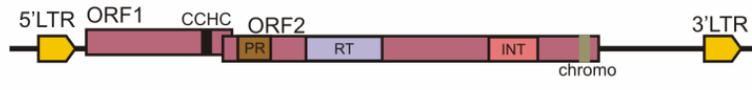
Pc_Metavir4
(7615 bp)



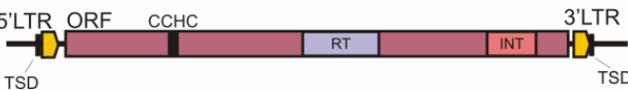
Pc_Metavir5
(7635 bp)



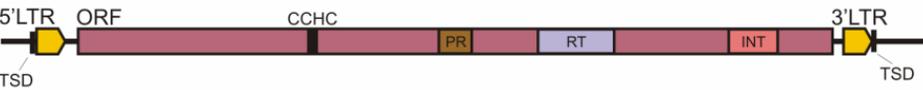
Pc_Metavir6
(8003 bp)



Pc_Metavir7
(6024 bp)



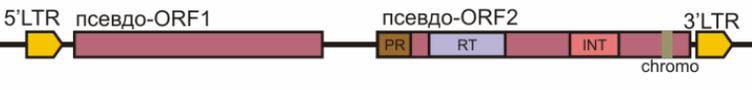
Pc_Metavir8
(9326 bp)



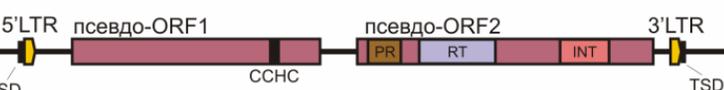
Pc_Metavir9
(8843 bp)



Pc_Metavir10
(7986 bp)

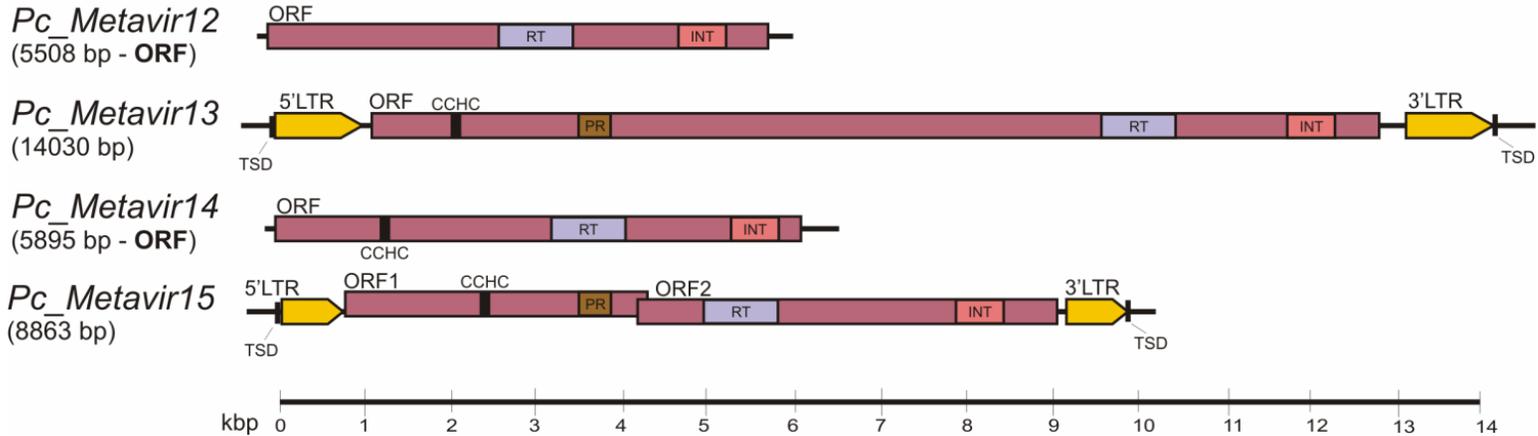


Pc_Metavir11
(7601 bp)



ORF - открытая рамка считывания; **5'LTR** и **3'LTR** - 5' и 3' длинные концевые повторы; **RT** - домен обратной транскриптазы; **Int** - интегразы; **PR** - протеаза; **dUTPase** - катализирует гидролиз dUTP; **chromo** - хромодомен; **CCHC** - цистеиновый мотив; * - стоп-кодон; **TSD** - повторы-мишени.

Metaviridae из *Phanerochaete chrysosporium*



ORF - открытая рамка считывания; **5'LTR** и **3'LTR** - 5' и 3' длинные концевые повторы; **RT** - домен обратной транскриптазы; **Int** - интегразы; **PR** - протеаза; **CCHC** - цистеиновый мотив; * - стоп-кодон; **TSD** - повторы-мишени.

Столь высокое разнообразие LTR ретротранспозонов в геноме одного вида грибов показано впервые.

В общей сложности LTR ретротранспозоны составляют около 3% от генома *P. chrysosporium*.

Выводы

1. Исследовано разнообразие non-LTR ретротранспозонов в геномах 22-х видов отряда Scorpiones. Показано присутствие трех филогенетических групп non-LTR ретротранспозонов (CR1, Jockey и I) и множественных линий внутри каждой из них.
2. Исследовано разнообразие non-LTR ретротранспозонов в геномах четырех видов бабочек рода *Maculinea*. Показано присутствие четырех филогенетических групп non-LTR ретротранспозонов (CR1, R1, Jockey и R4) и наличие множественных линий внутри них, в частности, двух близкородственных семейств MacCR1A и MacCR1B внутри филогенетической группы CR1.
3. Обнаружен *VmCR1B* элемент в геноме *Bombyx mori*, имеющий высокое сходство с *MacCR1B* элементом *Maculinea*.
4. Проведено исследование распределения CR1A и CR1B семейств среди Lepidoptera. Обнаружен горизонтальный перенос *CR1B* non-LTR ретротранспозонов между представителем семейства Bombycidae и общим предком рода *Maculinea* более 5 млн. лет назад.

Выводы

5. Исследовано разнообразие LTR ретротранспозонов в геномах восьми эукариотических организмов, включая три представителя царства грибов (Fungi) и пять представителей царства животных (Animalia). Было обнаружено 42 новых, ранее не описанных, LTR ретротранспозонов.
6. Проведен анализ геномов двух аскомицетов, *Aspergillus fumigatus* и *A. nidulans*. Для каждого вида было обнаружено по два семейства LTR ретротранспозонов, которые не были описаны ранее.
7. Проведенный RIP анализ убедительно показал, что последовательности LTR ретротранспозонов *A. fumigatus* и *A. nidulans* несут отпечатки действия RIP-подобной инактивации.
8. Исследовано разнообразие LTR ретротранспозонов в полной геномной последовательности лигнин-деградирующего базидиомицета *Phanerochaete chrysosporium*. Было обнаружено 8 новых Pseudoviridae и 15 элементов группы Metaviridae. Столь высокое разнообразие LTR ретротранспозонов в геноме одного вида грибов показано впервые. Большинство обнаруженных LTR ретротранспозонов *P. chrysosporium* представлены в геноме интактными полноразмерными копиями, которые относительно недавно перемещались. В общей сложности LTR ретротранспозоны составляют около 3% от генома *P. chrysosporium*.



Финансовая поддержка:

Программа Президиума РАН "Происхождение и эволюция биосферы" №10104-34/П-18/155-270/1105-06-001/28/2006-1

This study was financed by EC within its RTD project "MacMan" EVK2-CT-2001-00126.

