



Организация геномов эукариот: экспериментально-теоретические подходы к исследованию мобильных генетических элементов

Блинов Александр

В рамках вводного спецкурса «ИНФОРМАЦИОННАЯ БИОЛОГИЯ»

Что должен знать и уметь современный молекулярный биолог?

I.

Экспериментальная работа:
Умение спланировать эксперимент
Владение методами экспериментальной работы
Постановка эксперимента
Адекватная интерпретация результатов

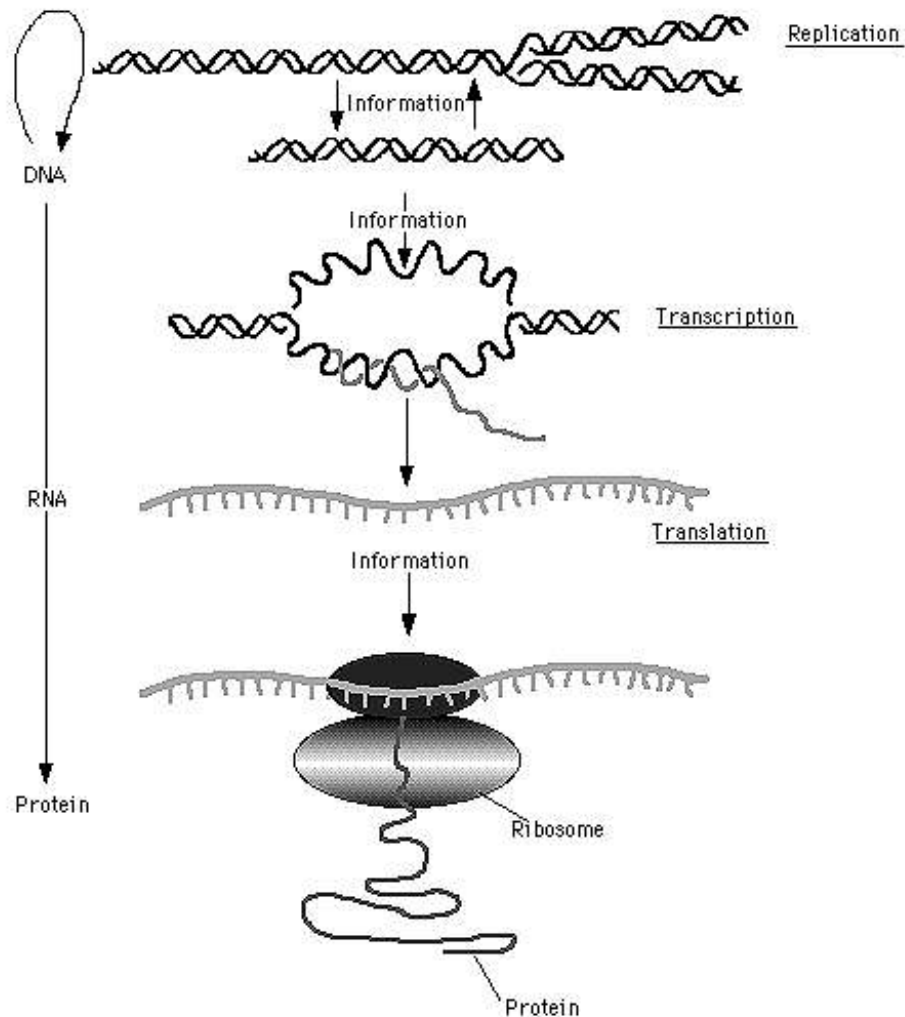
II.

Теоретическая работа:
Умение постановки задачи и выбора условий реализации решения
Работа с базами данных
Работа с различными программами для анализа последовательностей

Высококвалифицированный специалист



Фундаментальные генетические процессы, протекающие в эукариотической клетке.

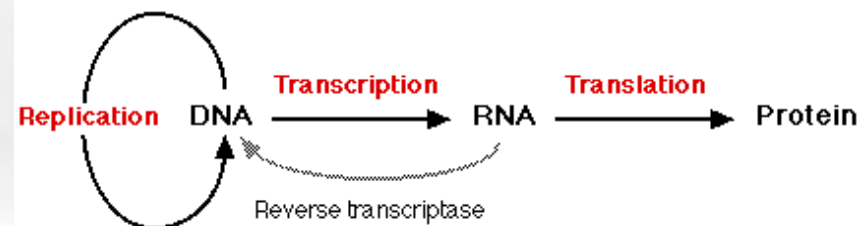


Репликация – удвоение ДНК

Транскрипция – синтез молекулы мРНК по матрице ДНК

Трансляция – синтез белка на рибосомах по матрице мРНК

Центральная догма молекулярной биологии



Что такое мобильные элементы?

Мобильные элементы (МЕ) – это последовательности ДНК, способные менять свое положение в геноме, то есть - перемещаться

Основная характеристика - многокопийность

Вид	Размер гаплоидного генома (Mbp)	МЕ в геноме (%)
<i>Homo sapiens</i>	3300	35
<i>Zea mays</i>	2400	>50
<i>Drosophila melanogaster</i>	140	14-18
<i>Arabidopsis thaliana</i>	97	5

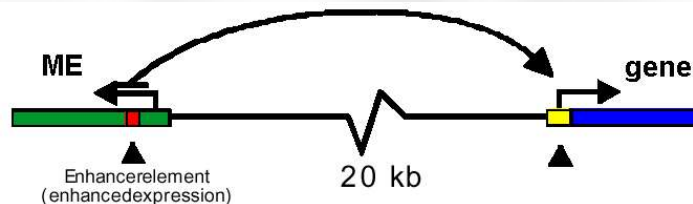
Взаимоотношение МЕ и генома хозяина

Первоначально предполагалось, что МЕ являются так называемой “эгоистичной ДНК” (selfish DNA). Однако, по мере накопления данных о взаимодействии МЕ и генома хозяина, эта точка зрения претерпевает изменения.

На сегодняшний день, данные о влиянии МЕ на геном хозяина остаются разрозненными

Известно, что:

- * Перемещаясь, МЕ изменяют структуру генома хозяина
- * Внутренние регуляторные части МЕ могут выступать как регуляторы экспрессии для близлежащих генов



* МЕ могут участвовать в образовании теломер

* И многое другое...

Классификация МЕ эукариот

На сегодняшний день не сформирована удовлетворяющая всех классификация мобильных элементов. Однако, существует разделение МЕ эукариот на два основных класса в соответствии с их структурной организацией и механизмом, который используется для перемещения

→ I класс РЕТРОТРАНСПОЗОНЫ

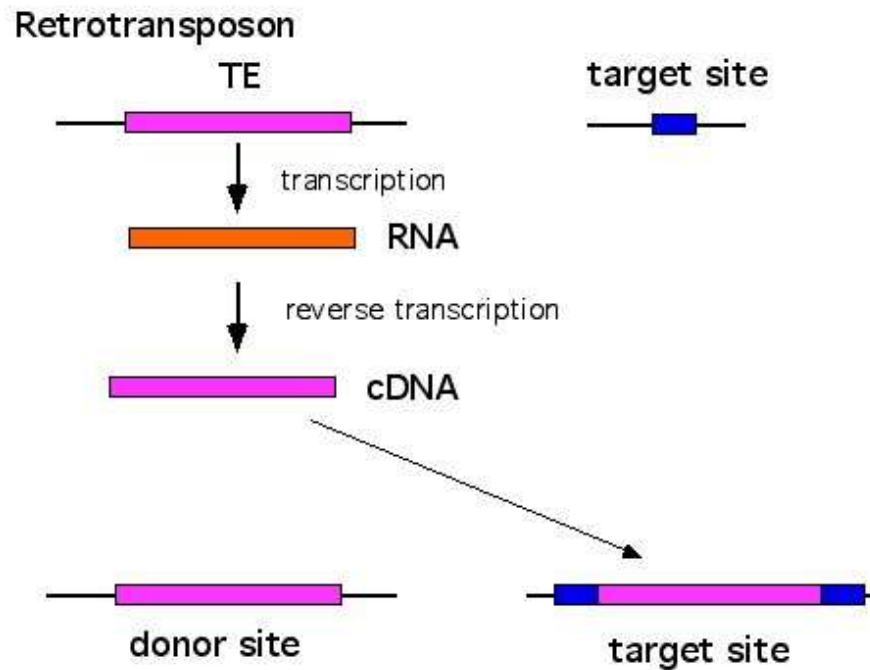
→ II класс ДНК ТРАНСПОЗОНЫ

→ III класс ОСТАЛЬНЫЕ

Механизмы перемещения эукариотических МЕ (1)

Ретротранспозиция (Retrotransposition)

→ Обобщенная схема



На первом этапе ретротранспозиции происходит считывания транскрипта, при последующей трансляции которого, синтезируется набор ферментов, необходимых для дальнейшего процесса ретротранспозиции.

На следующем этапе происходит синтез кДНК, ключевым ферментом является обратная транскриптаза (*reverse transcriptase* – отсюда и название “*retrotransposition*”).

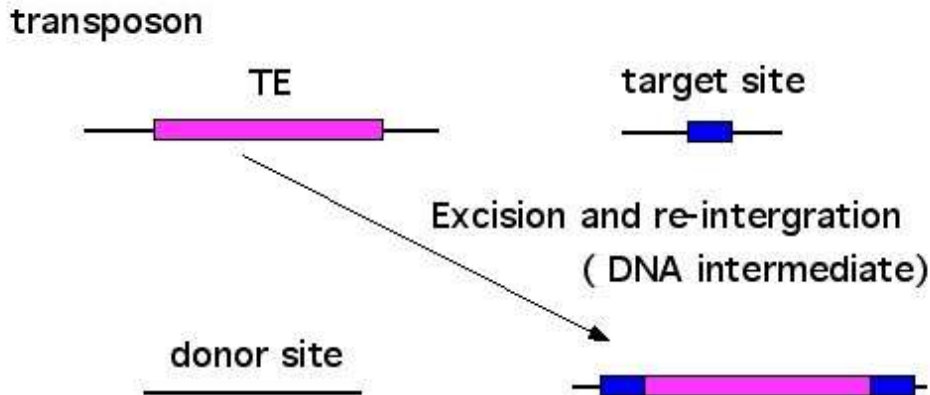
На последнем этапе происходит встраивание вновь образовавшейся копии элемента в геном.

→ Приведенная схема ретротранспозиции представляет собой обобщение. В действительности, существует как минимум три механизма ретротранспозиции. В рамках данного курса останавливаться на каждой из них не представляется возможным.

Механизмы перемещения эукариотических МЕ (2)

ДНК транспозиция или консервативная транспозиция

Этот механизм перемещения называется консервативной транспозицией, потому что в его процессе не происходит образование новых копий элемента.



В процессе консервативной транспозиции происходит вырезание одной из копий элемента и встраивание ее в другой участок генома.

Увеличение числа копий элементов, перемещающихся таким путем, происходит, главным образом, в процессе рекомбинаций.

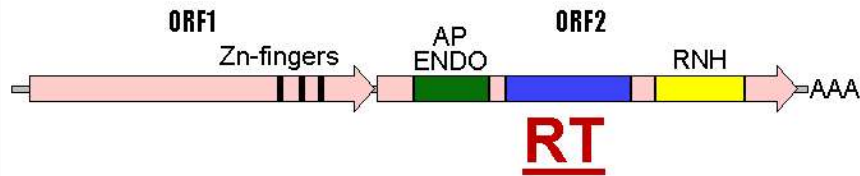
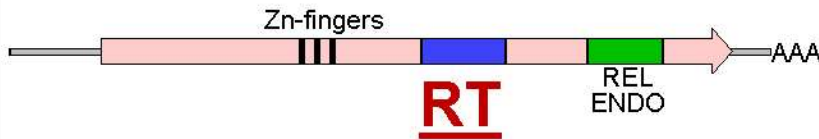
Еще одно название ДНК транспозиции – механизм “вырезания-встраивания” (cut-and-paste)

Структурная организация различных групп МЕ эукариот (1)

I. Ретротранспозоны

1. Ретропозоны

non-LTR ретротранспозоны



SINE элементы



Ретропозоны, в отличие от LTR ретротранспозонов, которые будут рассмотрены ниже, не несут на концах длинных прямых повторов.

Среди ретропозонов выделяют: non-LTR ретротранспозоны и SINE элементы.

Non-LTR ретротранспозоны имеют одну или две открытые рамки считывания (ORF) – обязательно присутствие последовательности, кодирующей обратную транскриптазу. На 3' конце элемента располагается *polyA* последовательность. Есть дополнительно кодируемые ферментативные домены – эндонуклеазы и рибонуклеаза H.

SINE элементы не несут последовательностей, кодирующих какой-либо белок, однако, в их структуре можно выделить следующие части: *pol III* промотор – чаще всего похож на *tRNA pol III* промотор; коровая или центральная часть; *polyA* последовательность на 3' конце, которая имеет высокую гомологию с 3' *polyA* последовательностями non-LTR ретротранспозоны из того же генома.

Замечание #1:

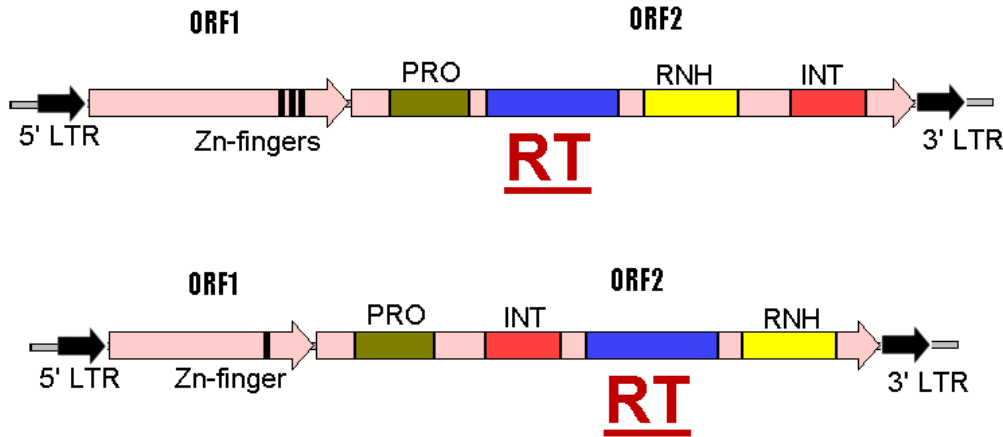
SINE элементы, зачастую, являются видо- или родо-специфичными, это навело на мысль об образовании *SINE* элементов *de novo*. Однако, механизм этого процесса до сих пор неизвестен.

Замечание #2:

Все вышесказанное относится к полноразмерным копиям non-LTR ретротранспозонов, однако, большинство копий в геноме, чаще всего, нарушены и могут не иметь многие из перечисленных характеристик.

Структурная организация различных групп МЕ эукариот (2)

2. LTR ретротранспозоны



По своей структуре LTR ретротранспозоны гораздо более разнообразны, чем ретропозоны. общей чертой всех LTR ретротранспозонов является наличие длинных концевых повторов (LTRs). Длинные концевые повторы несут регуляторные последовательности, в первую очередь – внутренний промотор (5' LTR), а также сигнал полиаденилирования (3' LTR).

Количество открытых рамок считывания может варьировать от одной до трех, причем, первая из них кодирует ДНК-связывающий белок, вторая – необходимые для ретротранспозиции ферменты, третья - так называемый белок оболочки (envelope), в результате, LTR ретротранспозон может обладать инфекционной способностью, как ретровирусы. Известны случаи наличия дополнительных рамок и/или антисенсе-рамок, кодирующих ферменты, которые не необходимы для ретротранспозиции.

Замечание #1:

Количеством рамок, разнообразие структуры LTR ретротранспозонов не ограничивается. По структуре второй рамки, которая кодирует большинство ферментов, необходимых для ретротранспозиции, LTR ретротранспозоны могут быть четко разделены на несколько групп.

Замечание #2:

Одна из групп LTR ретротранспозонов, как считается на сегодняшний день, дала начало ретровирусам, к которым относятся, например, вирус иммунодефицита человека (HIV). Об этом будет рассказано далее...

Замечание #3:

Как и в случае с non-LTR ретротранспозонами, большая часть копий LTR ретротранспозонов в геноме может быть нарушена.

Структурная организация различных групп МЕ эукариот (3)

II. ДНК транспозоны



Замечание #1:

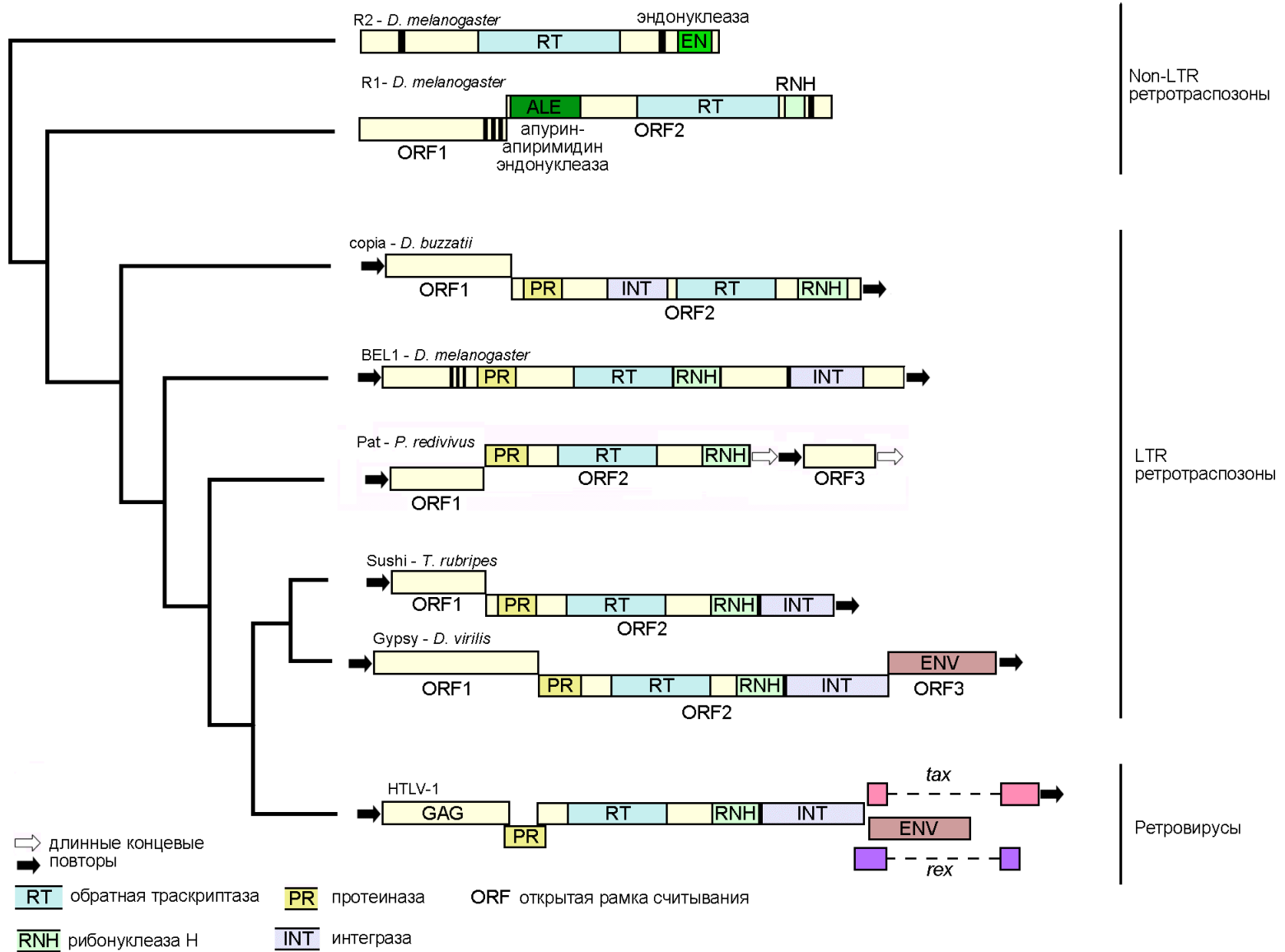
Именно ДНК транспозоны были первыми мобильными элементами, обнаруженными у эукариот. Первооткрывательница Барбара МакКлинток получила в 1983 году Нобелевскую премию

ДНК транспозоны характеризуются наличием инвертированных фланкирующих повторов и присутствием одной открытой рамки считывания, кодирующей транспозазу, ключевой фермент процесса транспозиции по механизму вырезания-встраивания.

Многие из ДНК транспозонов могут нести дополнительные открытые рамки или даже гены, которые захватываются из генома хозяина и не играют никакой роли в механизме транспозиции.

Значительную фракцию МЕ в геноме могут составлять нарушенные копии ДНК транспозонов, которые относятся к неавтономным ДНК транспозонам. Для своей транспозиции эти нарушенные МЕ могут использовать транспозазу, кодируемую автономным ДНК транспозоном.

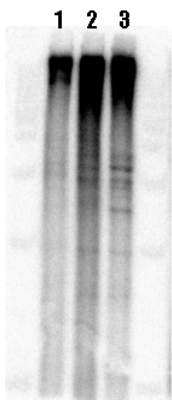
Происхождение различных групп ретротранспозонов



Основные принципы выявления МЕ (1)

Экспериментальные подходы

Первоначально, МЕ обнаруживали случайно. Классический пример, ДНК транспозоны были обнаружены Барбарой МакКлинток, когда она изучала изменчивость окраски зерен кукурузы.

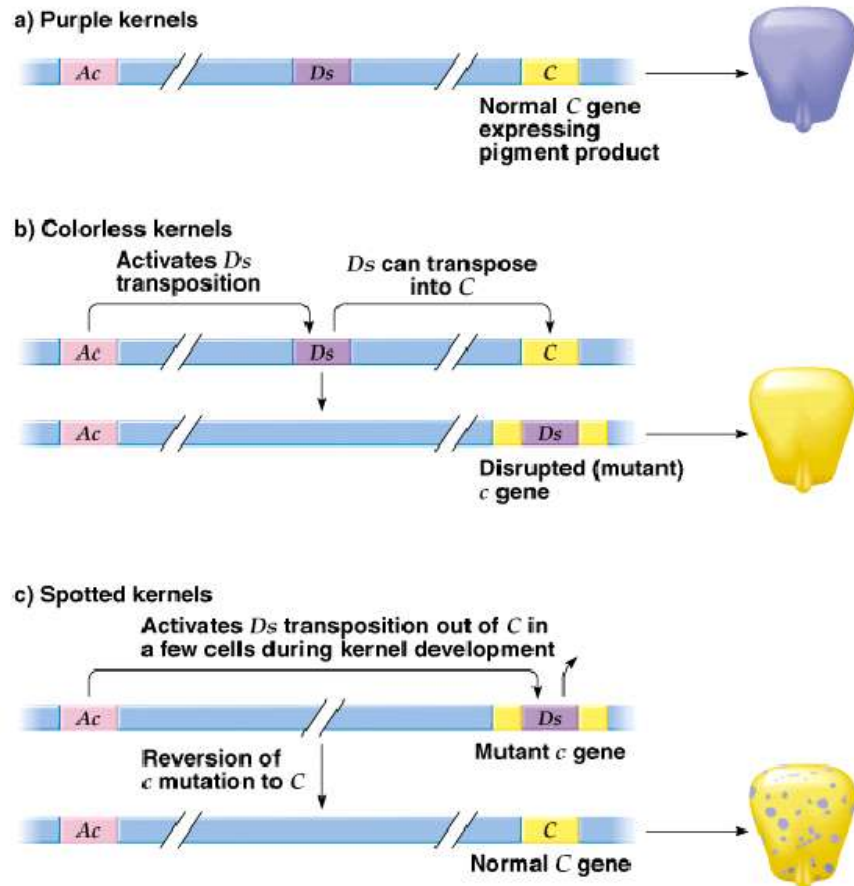


Саузерн-блот гибридизация на уже известные мобильные элементы.



Анализ при помощи полимеразной цепной реакции (PCR)

с использованием вырожденных праймеров с последующим клонированием и установлением первичной нуклеотидной последовательности клонов



Основные принципы выявления МЕ (2)

Биоинформатические подходы

1. Исторически, первым был реализован поиск МЕ в базах данных последовательностей при помощи **BLAST (Basic local alignment search tool)**.



Nucleotide

- Quickly search for highly similar sequences (megablast)
- Quickly search for divergent sequences (discontiguous megablast)
- Nucleotide-nucleotide BLAST (blastn)
- Search for short, nearly exact matches
- Search trace archives with megablast or discontiguous megablast

Protein

- Protein-protein BLAST (blastp)
- PHI- and PSI-BLAST
- Search for short, nearly exact matches
- Search the conserved domain database (rpsblast)
- Search by domain architecture (cdart)

Translated

- Translated query vs. protein database (blastx)
- Protein query vs. translated database (tblastn)
- Translated query vs. translated database (tblastx)

Genomes

- Chicken, cow, pig, dog, sheep, cat
- Environmental samples
- Human, mouse, rat
- Fugu rubripes, zebrafish
- Insects, nematodes, plants, fungi, malaria
- Microbial genomes, other eukaryotic genomes

В качестве запроса используются последовательности уже известных мобильных элементов. Понятно, что вероятность обнаружить, принципиально отличные от запроса МЕ, ничтожно мала. С другой стороны, это наиболее быстрый способ извлечения МЕ из интересующей базы данных.



На сегодняшний день NCBI является наиболее полной базой как нуклеотидных, так и аминокислотных последовательностей. Кроме того, предлагает все возможные варианты BLAST, разработанные на данный момент.

National Center for
Biotechnology Information

Основные принципы выявления МЕ (3)

2. Поиск **всех повторов** в целом геноме или некоторой последовательности при помощи специальных программ, с последующим вычленением МЕ.



RepeatMasker

Одной из таких программ является **RepeatMasker**.
Большим недостатком этой программы является то, что она доступна только в виде Web интерфейса.



Институт GIRI дает возможность поиска МЕ при помощи программы **CENSOR** по e-mail запросу.

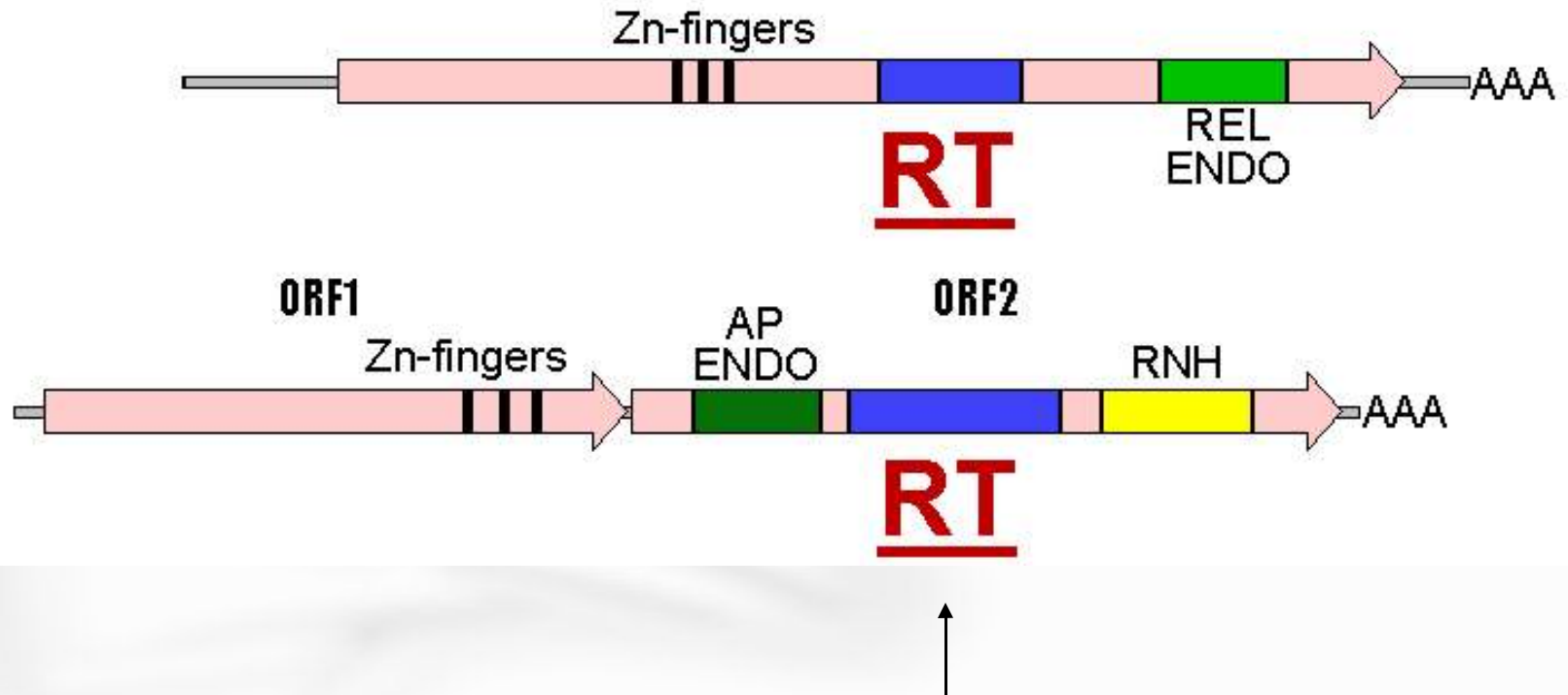
CENSOR и **RepeatMasker** разработаны одними и теми же авторами и используют единую базу данных повторенных последовательностей. По большому счету, это два названия одного и того же метода.

3. Наиболее интересным на данный момент представляется **целенаправленный поиск** МЕ, принадлежащих к той или иной группе по характеристикам данной конкретной группы.
Шаги в этом направлении предпринимаются сегодня, в частности была разработана программа для поиска LTR ретротраспозонов по длинным концевым повторам – **LTR_STRUC**. Для поиска отдельных групп элементов применялись алгоритмы на основе **цепей Маркова**.

Однако, пока не разработано методов для целенаправленного поиска SINE элементов, полуманных элементов

Принципы целенаправленного поиска различных групп МЕ (1)

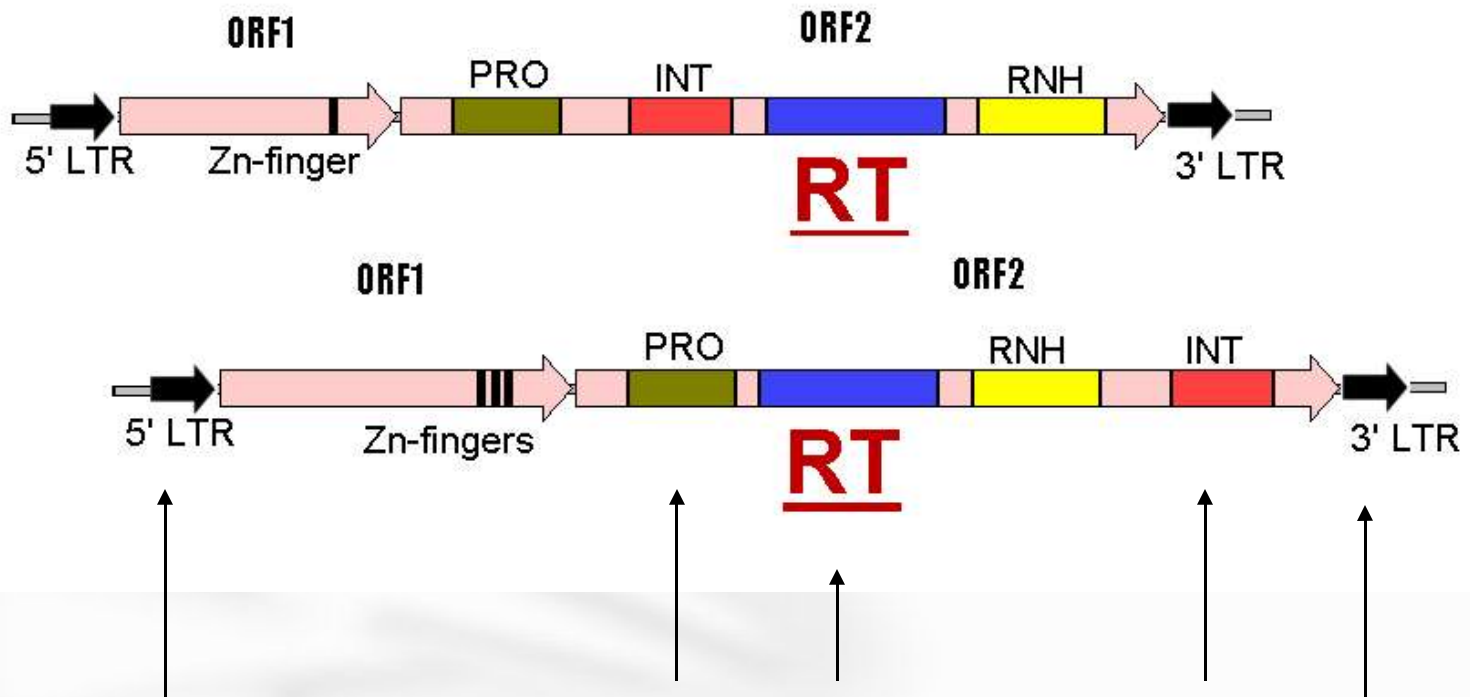
I. Non-LTR ретротранспозоны



Обратная транскриптаза (RT) – единственный домен, присутствующий у всех non-LTR ретротранспозонов. Поиск на обратную транскриптазу известных non-LTR ретротранспозонов позволяет обнаруживать non-LTR ретротранспозоны, например, в прочитанных геномах. Однако, эта стратегия не позволяет выявлять поломанные копии элементов, которые уже не содержат домен обратной транскриптазы в том виде, чтобы она могла быть обнаружена при помощи BLAST, алгоритмов на основе цепей Маркова и др.

Принципы целенаправленного поиска различных групп МЕ (2)

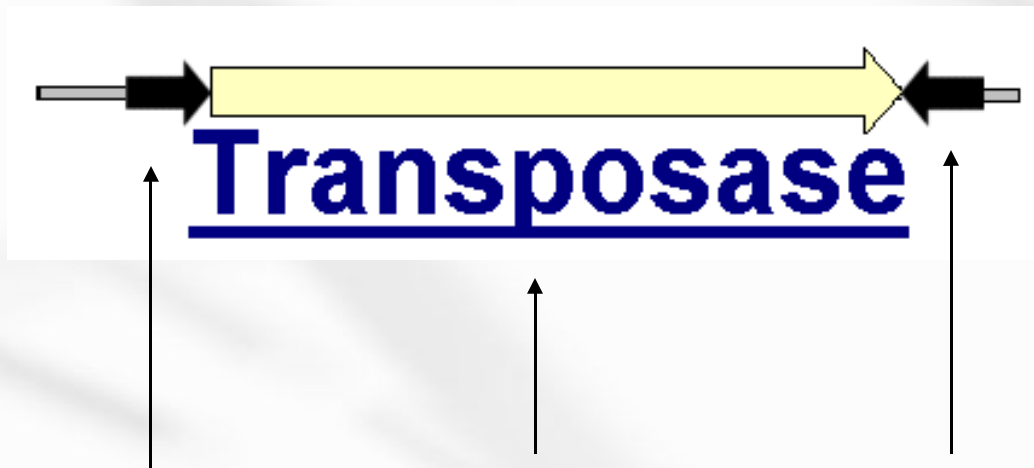
II. LTR ретротранспозоны



Как и в случае с non-LTR ретротранспозонами, поиск LTR ретротранспозонов может проводиться на домен обратной транскриптазы (RT), кроме того, non-LTR и LTR ретротранспозоны можно выявлять на этот домен совместно. Дополнительной характеристикой для поиска LTR ретротранспозонов могут служить длинные концевые повторы (LTR), а также наличие некоторых доменов, не свойственных другим МЕ, например – домены интегразы (INT) и протеиназы (PRO).

Принципы целенаправленного поиска различных групп МЕ (3)

III. ДНК транспозоны



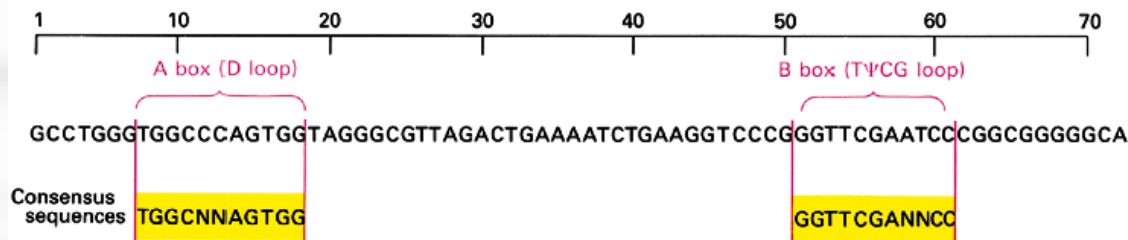
Транспозаза – домен, на который возможен поиск ДНК транспозонов, кроме того, инвертированные концевые повторы, являясь характеристикой ДНК транспозонов, могут быть использованы не только для поиска полных копий, но и для выявления нарушенных ДНК транспозонов в целом геноме или части последовательности.

Принципы целенаправленного поиска различных групп МЕ (4)

IV. SINE элементы



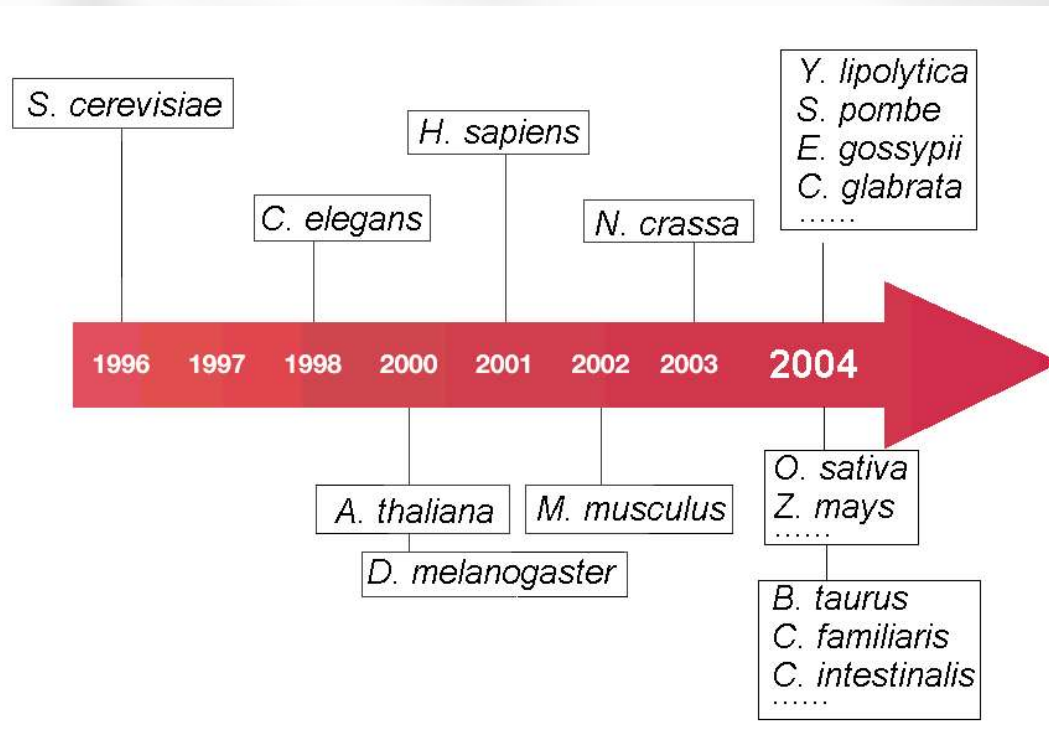
На сегодняшний день не существует алгоритмов, позволяющих вести целенаправленный поиск SINE элементов. В первую очередь это связано с тем, что SINE элементы не несут какие бы то ни было кодирующие участки. Единственное, что объединяет все SINE элементы – наличие pol III промотора в 5' части элемента. Один из методов, который применяется сегодня, для выявления SINE элементов заключается в поиске всех повторенных последовательностей, с последующим вычленением ДНК тарнспозонов, non-LTR и LTR ретротранспозонов. Оставшаяся фракция анализируется “в ручную”



Большинство известных на сегодняшний день SINE элементов имеют pol III промотор tRNA, который состоит из, так называемых, А и В боксов.

Предполагается, что на А и В боксы можно вести поиск SINE элементов, при задании дополнительных параметров, например – наличие polyA последовательности на 3' конце. Копийность SINE элементов, чаще всего, на порядки выше, чем у других МЕ, более того, SINE элементы, обычно, высокогетерогенны, что может облегчить их целенаправленный поиск.

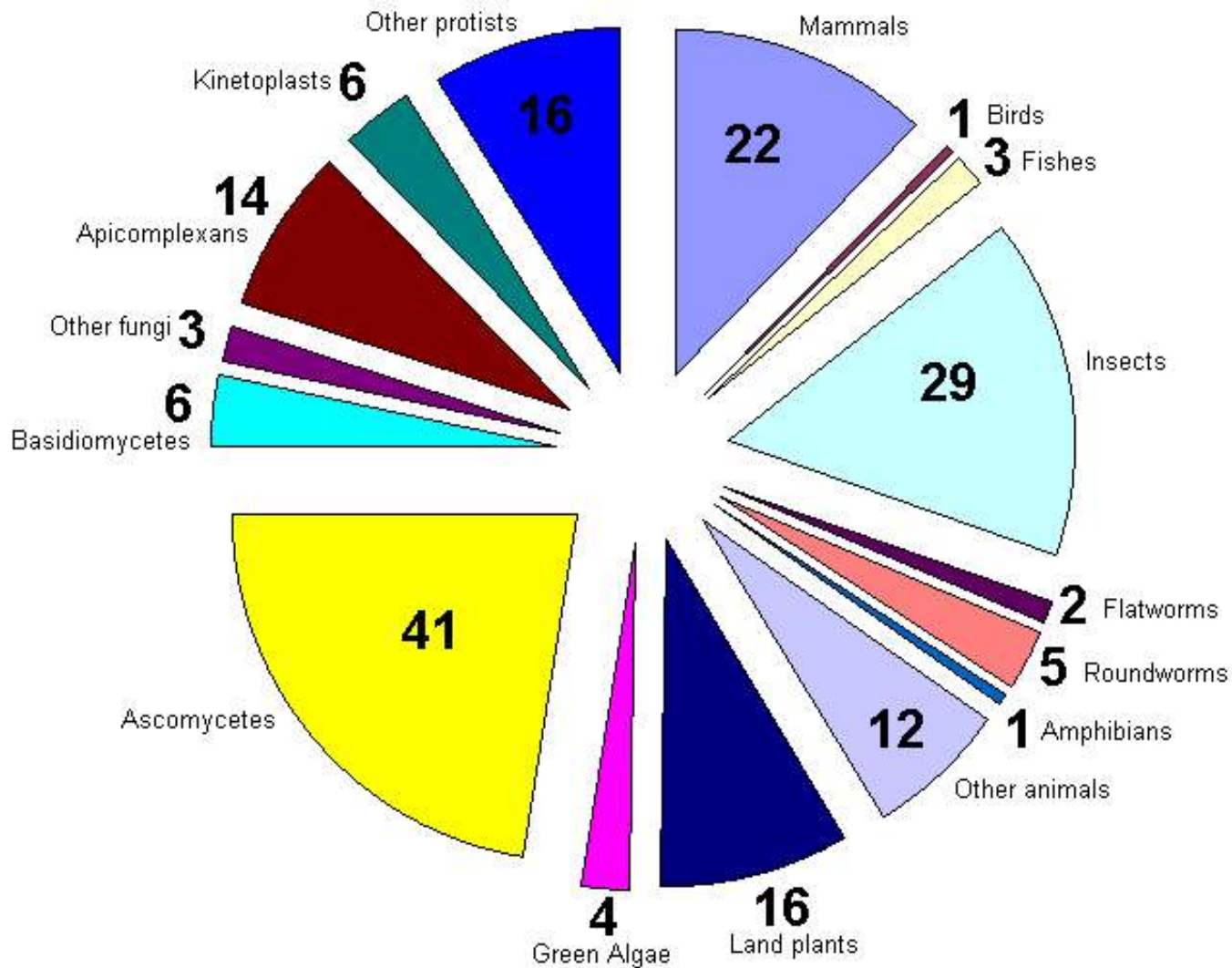
Почему необходим биоинформатический поиск МЕ?



За последние четыре года количество проектов по установлению полной последовательности геномов, в том числе и геномов эукариотических организмов, выросло на порядки. Более того, их количество неуклонно растет. Создание программ, при помощи которых будет возможен целенаправленный поиск МЕ может значительно облегчить проблемы аннотирования повторенных последовательностей и асемблирования геномов.

Еще несколько слов о ...

На сегодняшний день в базе NCBI зафиксировано **184 проекта** по установлению полной последовательности геномов эукариотических организмов, большинство из которых находятся в процессе реализации.



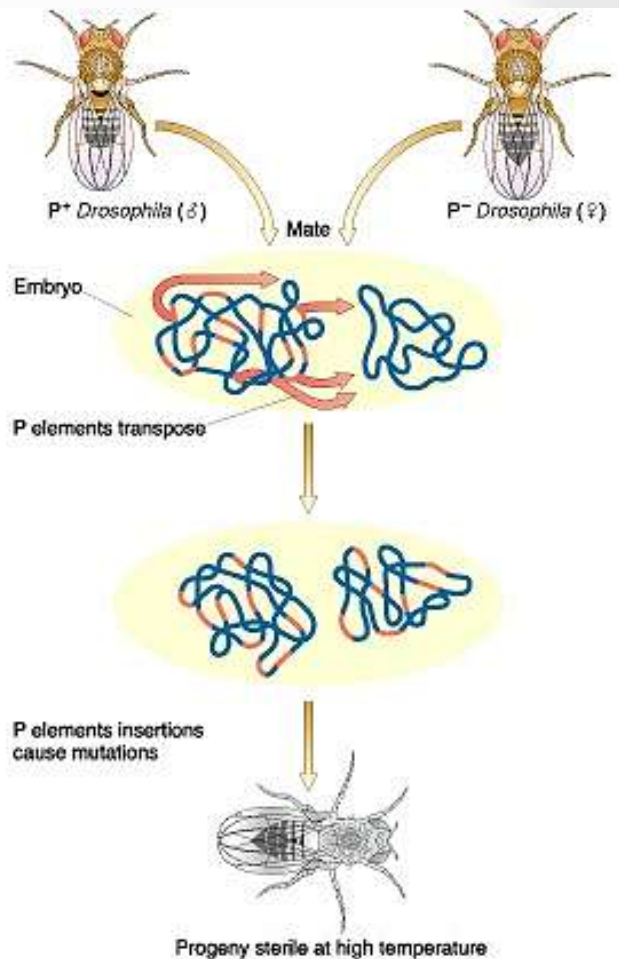
Всего же в базе NCBI зафиксировано **753 проекта** – подавляющее большинство из них направлены на установление последовательностей геномов бактерий.

Для чего необходимы исследования МЕ?

I.

МЕ являются многокопийными и могут составлять значительную часть генома, а значит влияние, которое МЕ могут оказывать на функционирование всего генома трудно переоценить. На сегодняшний день, мы обладаем разрозненными данными о влиянии МЕ на функционирование тех или иных генов, однако, накопление этих данных приведет к более глубокому пониманию не только процессов, происходящих в конкретном геноме, но и эволюции геномов.

Интересный факт: увеличение размера генома неразрывно связано с увеличением количества МЕ.



Наиболее ярким примером влияния МЕ на геном является гибридный дисгенез у *Drosophila*

Известны случаи заболеваний у человека, вызванные перемещением МЕ

Несколько примеров участия МЕ в формировании геномов:

Теломераза – фермент необходимый для формирования теломер, показано, что теломераза филогенетически связана с ретротранспозонами;

Так называемая V(D)J рекомбинация, которая необходима для формирования рецепторов иммунокомпетентных клеток происходит по механизму вырезания-встраивания, что говорит о филогенетической связи с транспозонами.

Для чего необходимы исследования МЕ?

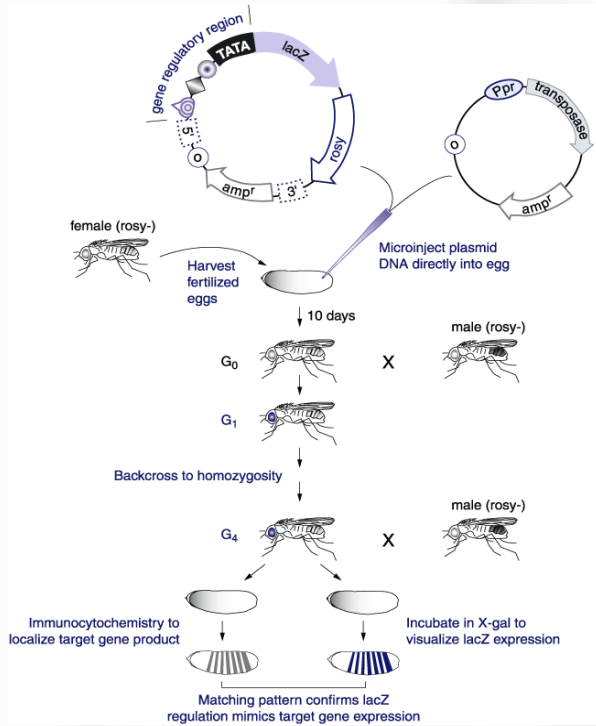
II.

МЕ имеют “практическое применение”:

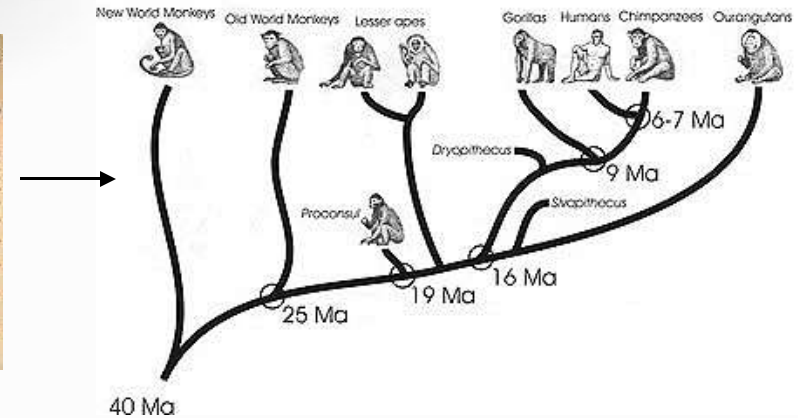
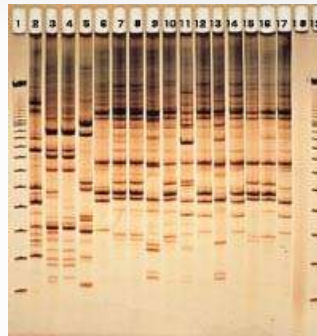
МЕ используются для получения трансгенных организмов.

МЕ используются для обнаружения и извлечения из генома новых генов, так называемый insertional mutagenesis – особенно активно применяется на растениях.

Все больше появляется информации о применении МЕ при генной терапии для лечения заболеваний.



МЕ могут использоваться как филогенетические маркеры, для построения эволюционных деревьев.



Так что же должен знать и уметь современный молекулярный биолог?!

I.

Экспериментальная работа:

Умение спланировать эксперимент

Владение методами экспериментальной работы

Постановка эксперимента

Адекватная интерпретация результатов

II.

Теоретическая работа:

Умение постановки задачи и выбора условий

реализации решения

Работа с базами данных

Работа с различными программами для анализа последовательностей



КАФЕДРА
ИНФОРМАЦИОННОЙ БИОЛОГИИ ФЕН НГУ



Лаборатория
теоретической генетики

Институт Цитологии и Генетики СО РАН

Высококвалифицированный специалист

