

# Поиск ретротранспозонов в больших последовательностях



**Институт  
Биоинформатики**

**Шульга Ольга  
Добрынин Павел**

Центр геномной биоинформатики им.Ф.Г.Добржанского, 2014г

# Актуальность работы

- Части генома эукариот являются повторяющимися.
- Основная доля повторов является мобильными элементами. Считается, что они оказывают значительное влияние на эволюцию гена и структуры хромосом.
- Мобильные элементы составляют около 50% генома позвоночных, и более высокий процент у видов растений.

# Введение

Группы мобильных элементов:

- **LTR ретротранспозоны**
- не LTR ретротранспозоны
- ДНК - транспозоны

# Введение

Методы распознавания ретровирусной последовательности:

- акцент на LTR
- акцент на повторы
- акцент на консервативные мотивы
- акцент на открытые рамки считывания

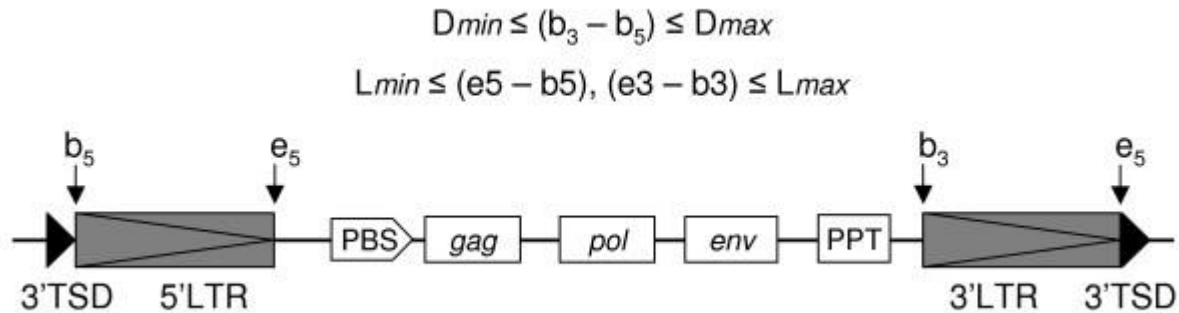
# Введение

Рассмотрены следующие ПО:

- LTR harvest
- LTR Finder
- LTR Struc
- RetroTector online

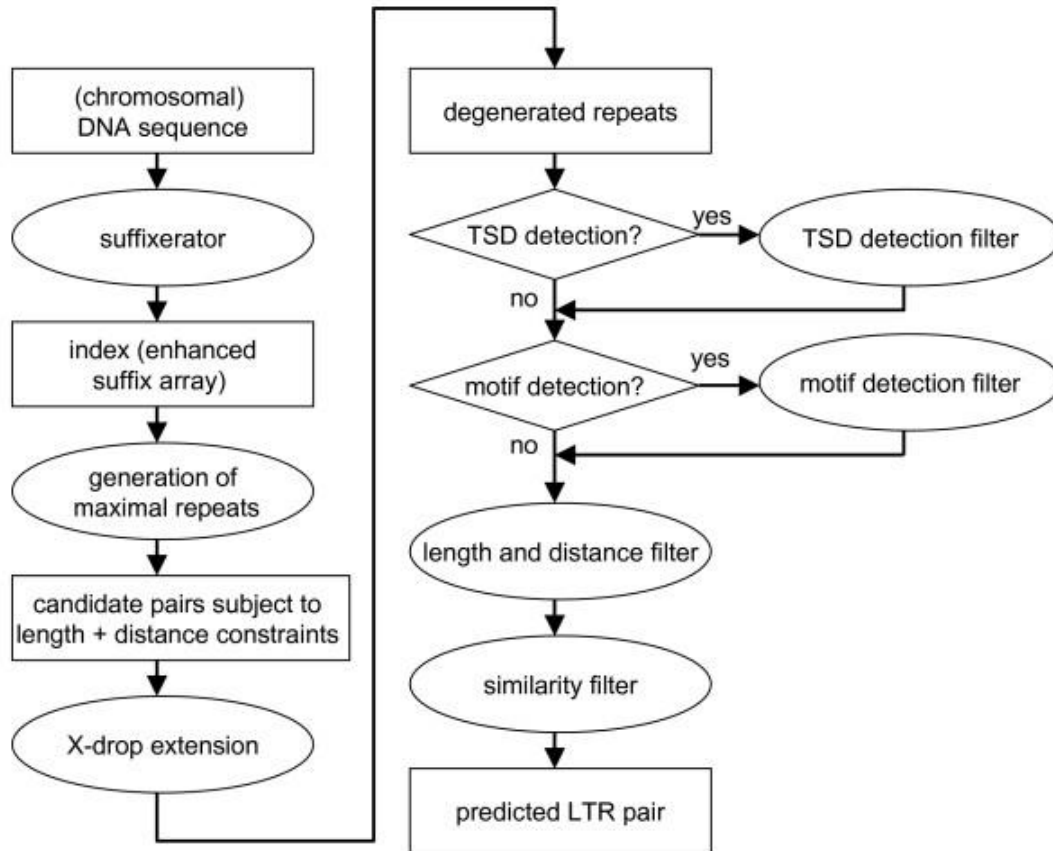


# LTR harvest



- Ограничение длины
- Ограничение диапазона
- Ограничение сходства
- TSD
- LTR motif

# LTR harvest



## Замечания:

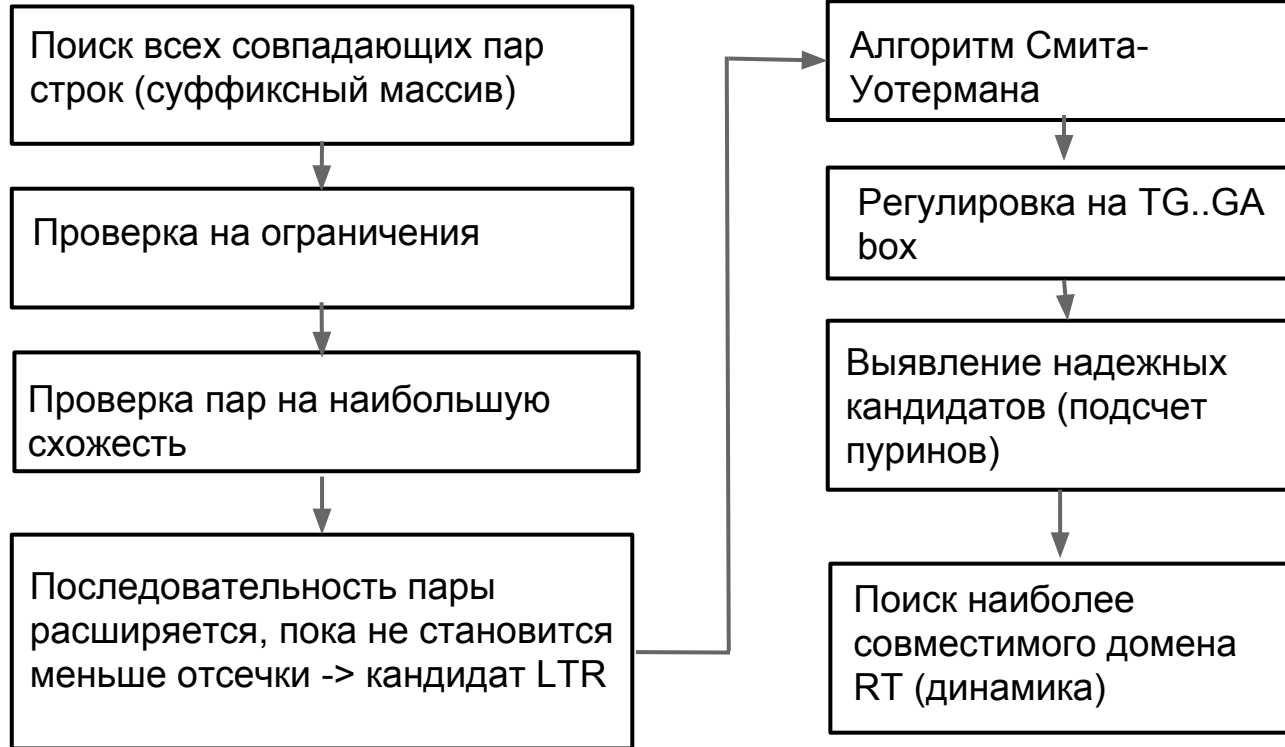
- Быстрота (~ 1-1.5 часа)
- Поиск в больших геномах
- Легкий в вычислениях шаг построения расширенного массива суффиксов (~ 1 час) (suffixerator)
- Поиск кандидатов LTR ретротранспозонов занимает несколько минут

# LTR harvest

- Происходят быстрые вычисления больших наборов данных (но до 50 000 000 байт )
- Гибкость настроек
- LTR harvest принимает последовательности в нескольких форматах
- Open source



# LTR finder



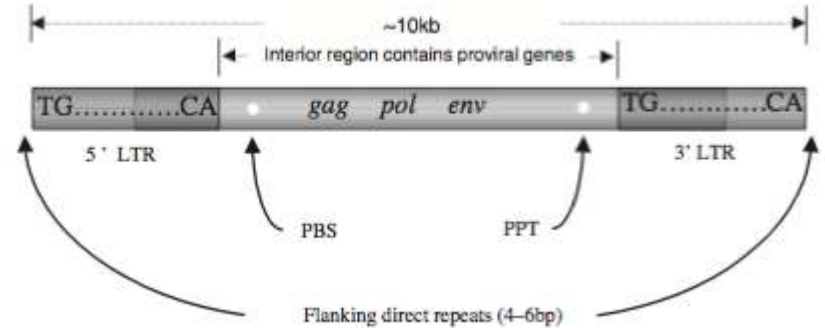
- Ограничение диапазона
- Ограничение длины
- Ограничение сходства

# LTR finder

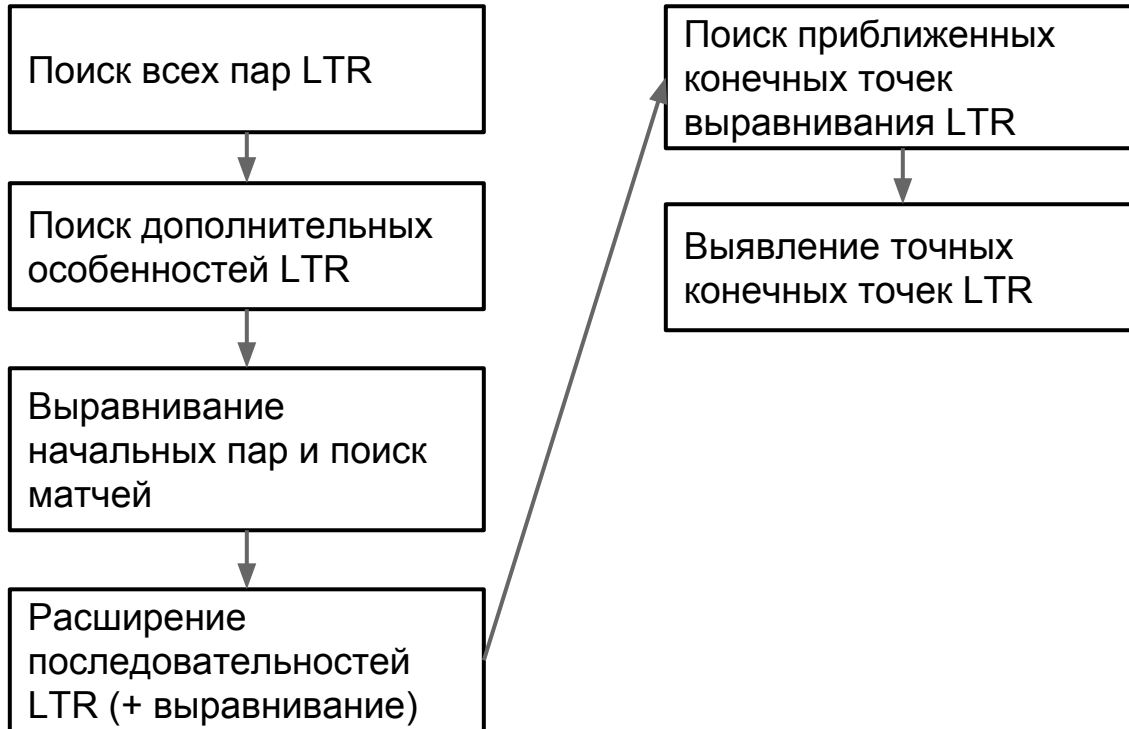
- Веб-сервис
- Эффективная обработка больших геномов (кукуруза, пшеница)
- LTR элементы, близкие к функциональным белкам представляются отдельно
- Ожидается очень много обновлений

# LTR struc

- Сайт связывания праймера PBS
- Полипурины PPT
- Наличие по крайней мере двунуклеотидных концов каждого LTR (обычно TG и CA)
- Прямой или “target-site” повторы



# LTR struc

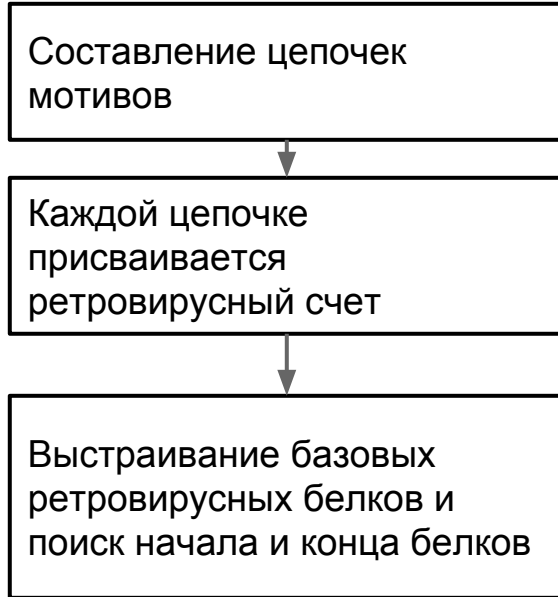


- Ограничение диапазона
- Ограничение длины
- Шаг выравнивания N (по умолчанию N=100)
- Уровень сходства 70%
- Рассчитывается мера вероятности реального положения ретротранспозона
- Язык программирования: Visual C++
- Только Fasta-формат
- Генерирует файл отчета с описанием

# RetroTector online

- “Цепной алгоритм”
- Испытание и оценивание коллекции консервативных мотивов
- Комбинированный алгоритм
- Модульное ПО
- Ориентирован на большие геномы позвоночных (входная последовательность до 10 Мб, лол )

# RetroTector online



- Не зависит от повторов
- Не зависит от одиночных LTR
- Не зависит от провирусов

# Вывод

Для поиска ретротранспозонов в больших данных подготавливает ПО LTR harvest. Другое ПО не прошло тест на поиск ретротранспозонов в больших геномах.