



# Сравнение программ RepeatMasker и LTRharvest

Научный руководитель: Павел Добрынин

Студент: Любовь Крайнова

Санкт-Петербург

2016

# Цели и задачи проекта:

**Цель:** сравнение двух подходов для аннотации повторов в геноме (на примере программ RepeatMasker и LTRharvest)

## **Задачи:**

- Научиться обрабатывать крупные файлы, используя команды языка Python;
- Познакомиться с принципами работы RepeatMasker и LTRharvest;
- Посмотреть разницу в результатах при обработке одного и того же файла двумя этими программами;
- На основании сравнения полученных результатов сделать выводы о различии в работе данных программ.

# Что такое RepeatMasker?

Ищет участки ДНК, содержащие повторы:

- Транспозоны
- Микро- и минисателлиты
- Участки со множественными повторами (Alu-повторы, AT или GC богатые регионы)
- Области ДНК низкой сложности

И «маскирует» их:

Hard-masking: **GGCGGCGGCGGCGGC** → NNNNNNNNNNN

Soft-masking: **AGCTTGGCGGCGGCGGC** → AGCTTggcggcggcggc



# Для чего используется RepeatMasker?

- BLAST-оптимизация
- Нахождение интересующих исследователя областей ДНК (вставки мобильных элементов, участки со множественными повторами);
- Подбор праймеров для участков с множественными повторами;
- Подбор праймеров, комплементарных хаотично расположенным участкам по всему геному.

# Принцип работы RepeatMasker:

- Поиск повторов в анализируемой последовательности;
- Сравнение найденных повторов с имеющимися в базе данных *Repbase* (repetitive DNA elements database)



Данные для исследования предоставлены:  
***Felis catus***: абиссинской кошкой Корицей



Первой домашней кошкой, для которой было сделано полногеномное секвенирование.

NCBI → Genome → assembly  
Felis\_catus\_8.0 (ID: 78)

# Получаемый на выходе файл:

3715 5504 (945)	3.7 0.2	0.4	scaffold0 1	1	458 (240379765) + L1_Fc	LINE/L1	5048
1279 6412 (10)	4.8 5.3	0.0	scaffold0 1	1258	1445 (240378778) + L1_Fc	LINE/L1	6215
613 224 (4)	22.7 7.9	4.0	scaffold0 2	1997	2212 (240378011) + SINEC1B_AMe	SINE/tRNA	1
566 201	19.4 6.7	11.4	scaffold0 3	2837	3046 (240377177) C SINEC1_AMe	SINE/tRNA	(0)
647 93 (436)	6.6 2.2	0.0	scaffold0 4	7010	7100 (240373123) - LTR1_FC	LTR/ERVk	1

Название транспозона

Семейство транспозона

Спасибо за внимание!

