

# Объединение моделей аннотации генов и коррекция аннотации

Полина Дроздова

Руководитель: Павел Добрынин, лаборатория геномной  
биоинформатики им. Ф.Г. Добржанского

30 мая 2015 г.

# Аннотация генома: общие сведения



Последовательность действий:

- 1 маскировка повторов (*de novo* или на основе гомологии);
- 2 выравнивание известных кодирующих последовательностей;
- 3 поиск генов *de novo*;
- 4 сведение аннотаций, полученных разными способами.

# Исходные данные и цель работы

## Предпосылки:

- 1 есть данные для недавно секвенированного штамма дрожжей *Saccharomyces cerevisiae*;
- 2 известно, что в нём есть гены, отсутствующие в геноме референсного штамма

## Цель работы:

Создание единой аннотации генома штамма 15В-П4 *Saccharomyces cerevisiae*.

## Задачи:

- 1 аннотация с помощью Augustus, GeneMark, Exonerate, litfOver, GlimmerHMM, SNAP;
- 2 сравнение аннотаций, сделанных разными алгоритмами;
- 3 сведение аннотаций и получение единого набора с помощью GLEAN, EVM, JIGSAW и/или Maker2.

# Использованные для аннотации инструменты

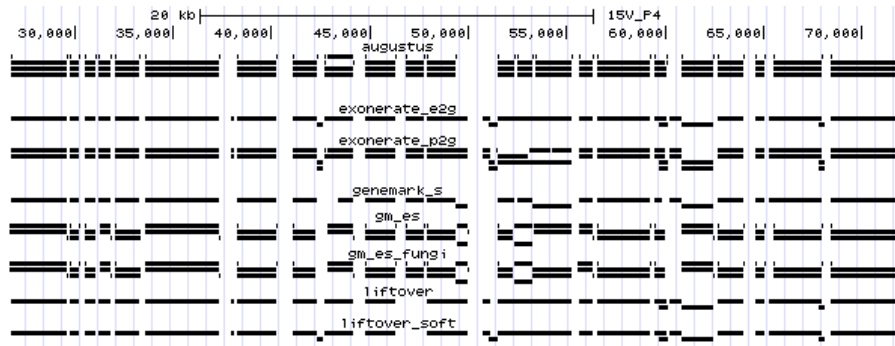
Инструменты, работающие на основе выравнивания:

- Exonerate 2.2.0
- UCSC liftOver (kentUtils 302)

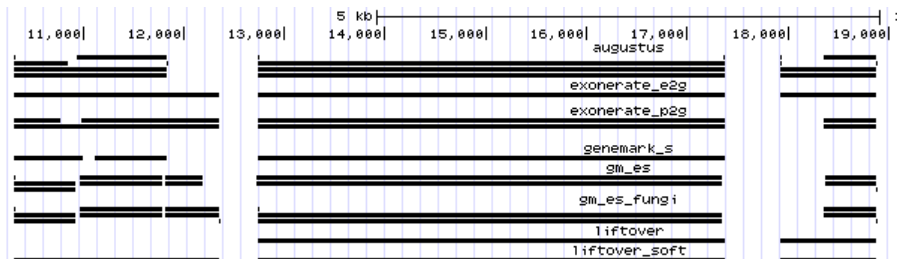
Предсказатели *de novo*:

- Augustus 3.0.3
- GeneMark-ES v4.21
- GeneMark-ES v4.21 –fungi
- GeneMark-S v4.30 –euk
- SNAP 2013-11-29

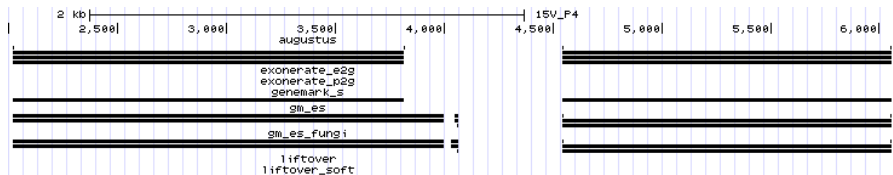
# Сравнение аннотаций в геномном браузере UCSC



## Иногда аннотации почти совпадают



# Иногда аннотации не совпадают



Гены из "винодельческого кластера"

## Сведение аннотаций

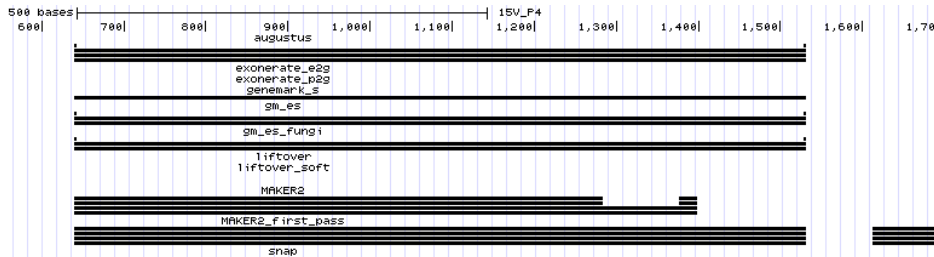
- JIGSAW ✗
- GLEAN ✗
- EVM ✗
- и/или Maker2 ✓



## Сведение аннотаций: Maker2

- JIGSAW X, проблема: не поддерживается, старый Perl
- GLEAN X, проблема: не запускается; автор советует Maker2
- EVM X, проблема: специфический формат файлов
- и/или Maker2.31.8 ✓

# Откуда MAKER берёт некоторые аннотации?



# Технические результаты

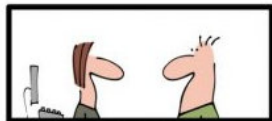
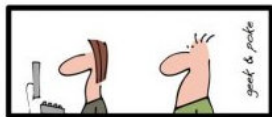
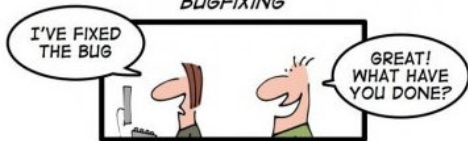
- 1 Аннотация с помощью:
  - Augustus, ✓
  - GeneMark, ✓
  - Exonerate, ✓
  - litfOver, ✓
  - GlimmerHMM,
  - SNAP. ✓
- 2 Сравнение аннотаций, сделанных разными алгоритмами:  
track hub в геномном браузере UCSC. ✓
- 3 Сведение аннотаций и получение единого набора с помощью:
  - GLEAN, ✗
  - EVM, ✗
  - JIGSAW ✗
  - и/или Maker2. ✓
- 4 Код для перевода разных форматов аннотации.

# Биологические результаты

- 1 Получена аннотация генов в геноме штамма 15В-П4.
- 2 В этом геноме есть некоторые гены, отсутствующие в референсном штамме; некоторые из них ранее найдены в других штаммах *S. cerevisiae*.
- 3 Впервые описан штамм, в геноме которого есть и «винодельческий кластер», и ген *RTM1*.

*THE ART OF  
BUGFIXING*

Спасибо за  
внимание!



*CHAPTER 1: SOMETIMES IT'S  
BETTER TO NOT EVEN TRY  
TO UNDERSTAND*