Перетасовка экзонов в геномах дрожжей: ошибка сборки или биологическая реальность

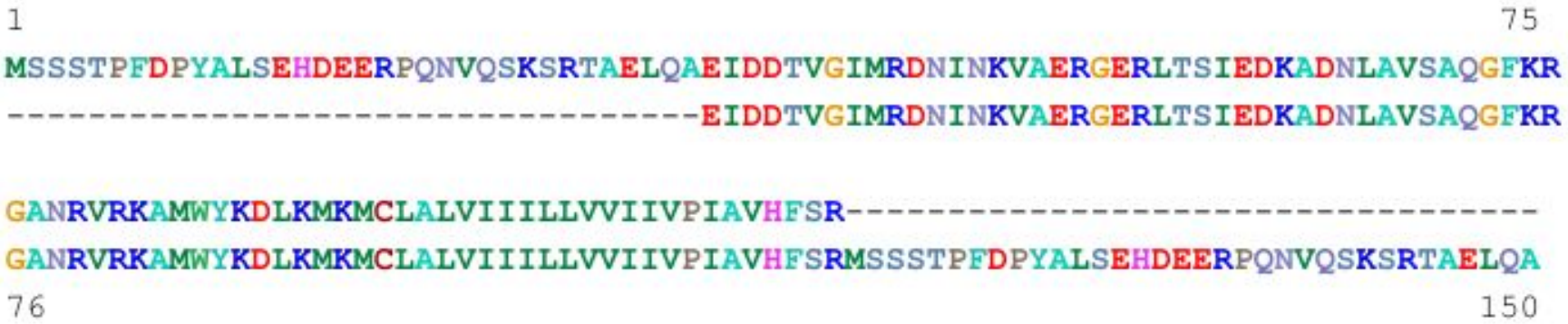
Выполнил: Раменский Дмитрий

Руководители: Олег Тарасов, Полина Дроздова
Кафедра генетики и биотехнологии СПбГУ

Предпосылки:

Ген *SNC1* (YAL030W) : Рецептор везикулярной мембраны (v-SNARE), участвует в слиянии транспортных пузырьков аппарата Гольджи с плазматической мембраной. Содержит два экзона и один интрон.

У части штаммов экзоны в аннотированных генах переставлены:



Задачи проекта:

1. Получить названия и последовательности всех генов *Saccharomyces cerevisiae* (возможно, и других видов *Saccharomyces* из базы yeastgenome.org), содержащих интроны.
2. Построить выравнивания по каждому из таких генов.
3. Проанализировать, есть ли другие гены (помимо *SNC1*) с перестановкой экзонов.
4. Пересобрать последовательности всех генов, в которых будут обнаружены перестановки экзонов, на основе первичных данных секвенирования.
5. Оценить возможные биологические последствия перестановки экзонов.

Задачи проекта:

1. **Получить названия и последовательности всех генов *Saccharomyces cerevisiae* (возможно, и других видов *Saccharomyces* из базы yeastgenome.org), содержащих интроны.**

В результате:

- С помощью поиска по базе данных http://yeastmine.yeastgenome.org/ получили 344 нуклеотидные последовательности *Saccharomyces cerevisiae*, содержащие интроны
- Из них 74 кодируют тРНК, мяРНК и митохондриальные белки - их убираем из рассмотрения

Задачи проекта:

2. **Построить выравнивания по каждому из таких генов.**

В результате:

Для построения выравниваний сгенерировали список ссылок на локусы интересующих нас генов в базе http://yeastgenome.org/ и использовали встроенную процедуру выравнивания между штаммами (Strain Alignment)

Задачи проекта:

3. **Проанализировать, есть ли другие гены (помимо *SNC1*) с перестановкой экзонов.**

В результате:

ORF Systematic Name	ORF Standard Name	ORF Name	ORF Length
YAL030W	*SNC1*	Suppressor of the Null allele of CAP	467
YBL050W	*SEC17*	SECretory	995
YBR255C-A			457
YCR097W	*HMRA1*	Hidden Mat Right A	487
YDL075W	*RPL31A*	Ribosomal Protein of the Large subunit	763
YER117W	*RPL23B*	Ribosomal Protein of the Large subunit	885
YFR045W			1002
YJL189W	*RPL39*	Ribosomal Protein of the Large subunit	542

Продолжение:

ORF Systematic Name	ORF Standard Name	ORF Name	ORF Length
YKL006C-A	*SFT1*	Suppressor of sed Five Ts	435
YLR329W	*REC102*	RECombination	892
YLR406C	*RPL31B*	Ribosomal Protein of the Large subunit	691
YLR445W	*GMC2*	Grand Meiotic recombination Cluster	649
YML034W	*SRC1*	Spliced mRNA and Cell cycle regulated gene	2631
YML036W	*CGI121*	homolog of human CGI-121	652
YNL004W	*HRB1*	Hypothetical RNA-Binding protein	1707
YNL038W	*GPI15*	GlycosylPhosphatidylInositol anchor biosynthesis	764

Задачи проекта

- **Оценить выравнивание белков с выявленными перестановками на геном**
4. **Пересобрать последовательности всех генов, в которых будут обнаружены перестановки экзонов, на основе первичных данных секвенирования.**

В результате:

Для выравнивания использовали программу Exonerate, позволяющую производить выравнивание белковых последовательностей на геном с учетом интронов

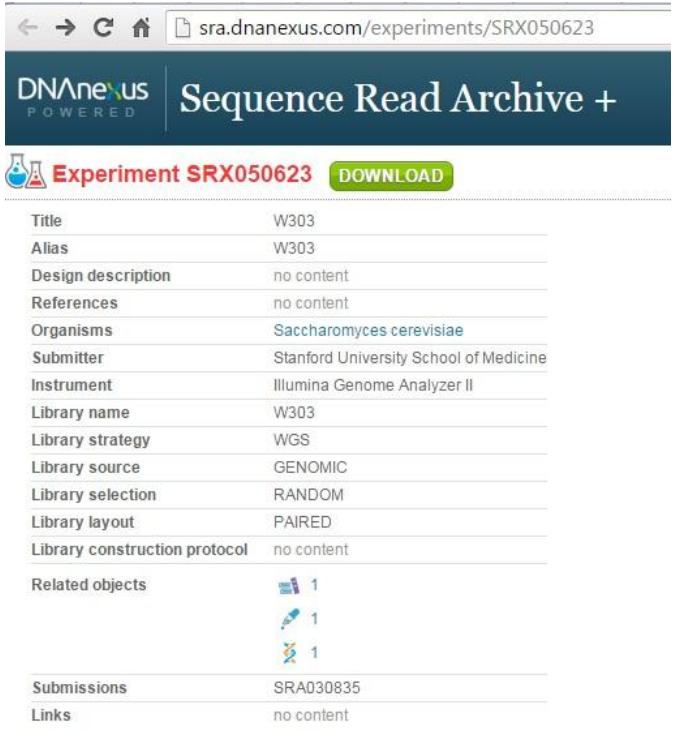
В результате:

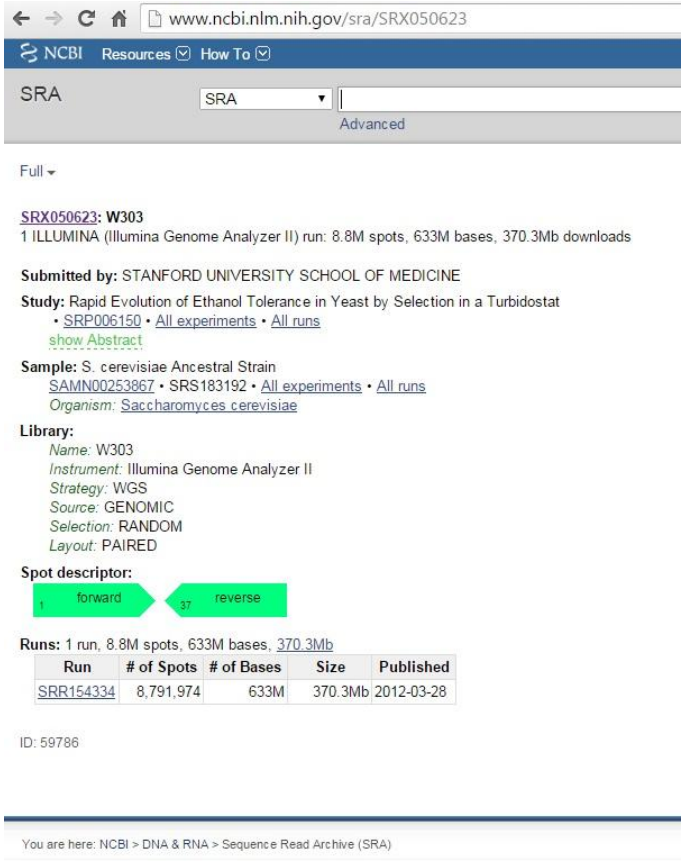
```
      1 : MetAlaArgAspIleThrPheLeuThrValPheLeuGluSerCysGlyAlaValAsnAsn :     20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          MetAlaArgAspIleThrPheLeuThrValPheLeuGluSerCysGlyAlaValAsnAsn
 155447 : ATGGCAAGAGATATCACATTTTTGACCGTATTTTTAGAAAGTTGTGGCGCTGTAAATA   : 155390

. . .

 40     : ProGluSerThrAspSerAsnSerLeuTyrIleProLeuLeuProPro{G}  >>>> Ta :     55
          ||||||||||||||||||||||||||||||||||||||||||||||||||{|}
          GluSerThrAspSerAsnSerLeuTyrIleProLeuLeuProPro{G}+-
 155332 :   GAATCAACCGACTCTAATTCATTATATATTCCACTGCTACCACCT{G}ga.......  : 155282

 56     : rget Intron 1 >>>>  {ly}MetLeuLysIleLysLeuAsnPheLysMetAsnAsp :     67
           97 bp              {||}!  :!!||||||||||||||||||||||||||||
                           ++{ly}LysValLysIleLysLeuAsnPheLysMetAsnAsp
 155281 : ..................ag{GA}AAAGTGAAGATTAAACTGAATTTTAAAATGAACG   : 155152
```

Задачи проекта

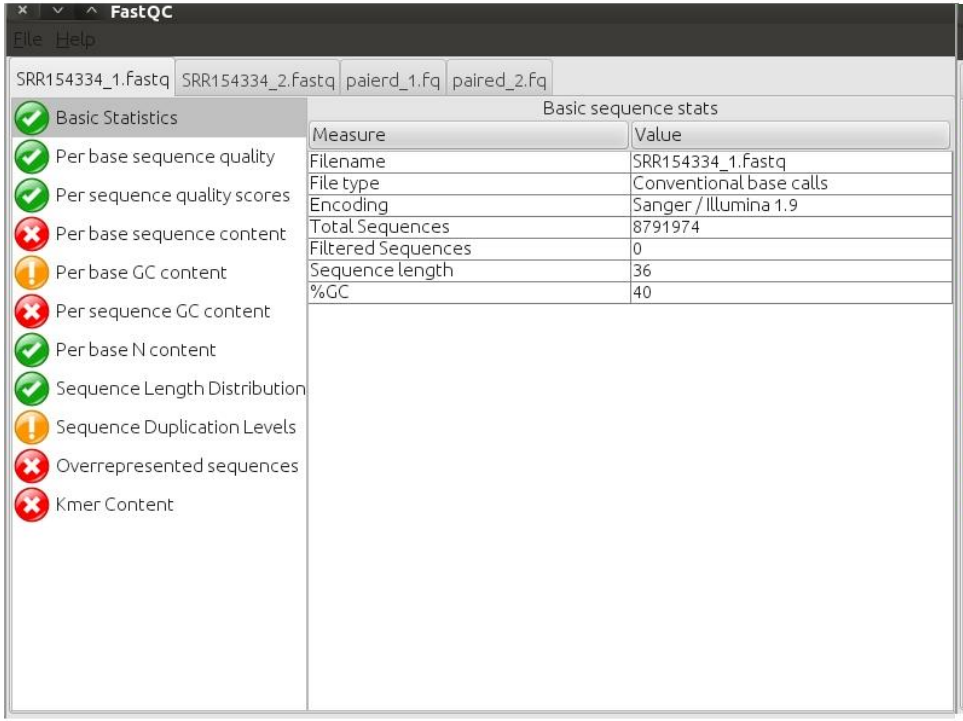
- **Оценить выравнивание белков с выявленными перестановками на геном**
4. **Пересобрать последовательности всех генов, в которых будут обнаружены перестановки экзонов, на основе первичных данных секвенирования.**

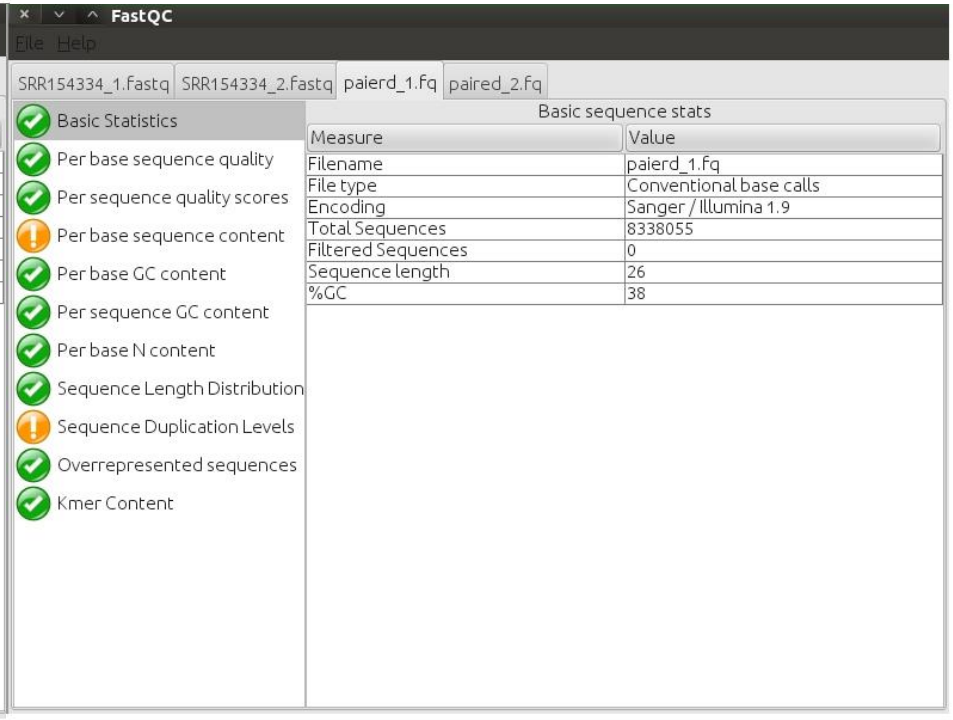
В результате:





В результате:





В результате: