

**Изучение полиморфизма генов устойчивости
класса NBS-LRR
пшеницы с использованием NGS,**

или

**Одна весьма поучительная история о
плохом проекте**

Анна Ненарокова

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М.В.ЛОМОНОСОВА

БИОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ
КАФЕДРА ГЕНЕТИКИ

**Анализ полиморфизма генов устойчивости класса NBS-LRR
в сортах пшеницы с использованием данных
высокопроизводительного секвенирования**

Дипломная работа студента V курса
Ненароковой Анны Александровны

Лаборатория системной биологии и
вычислительной генетики института
общей генетики им. Вавилова РАН

Научные руководители:
к.ф.-м.н. Касьянов Артём Сергеевич,
к.б.н. Пенин Алексей Александрович

Москва - 2015

Объект исследования: *Triticum* и *Aegilops*

Triticum aestivum



Иллюстрация с сайта
<http://www.agroatlas.ru>

Aegilops tauschii



Иллюстрация с сайта
<https://c2.staticflickr.com>

Объект исследования: гены устойчивости растений к патогенам



Гены устойчивости растений. Взаимодействие “ген-на-ген”

Генотип патогена	Генотип растения-хозяина			
	R_1-R_2-	$R_1-r_2r_2$	$r_1r_1R_2-$	$r_1r_1r_2r_2$
Avr_1-Avr_2-	устойчивый	устойчивый	устойчивый	восприимчивый
$Avr_1-avr_2avr_2$	устойчивый	устойчивый	восприимчивый	восприимчивый
$avr_1avr_1Avr_2-$	устойчивый	восприимчивый	устойчивый	восприимчивый
$avr_1avr_1avr_2avr_2$	восприимчивый	восприимчивый	восприимчивый	восприимчивый

R_n - ген устойчивости растения,

Avr_n - соответствующий ген авирулентности патогена

Классификация R-генов

1. PK

Серин/треонин-специфические цитоплазматические киназы

2. TM-LRR

Трансмембранные белки с доменом с обогащенными лейцином повторами

3. TM-LRR-PK

Трансмембранные белки с доменом с обогащенными лейцином повторами и серин/треонин-специфическим киназным доменом

4. TM-CC

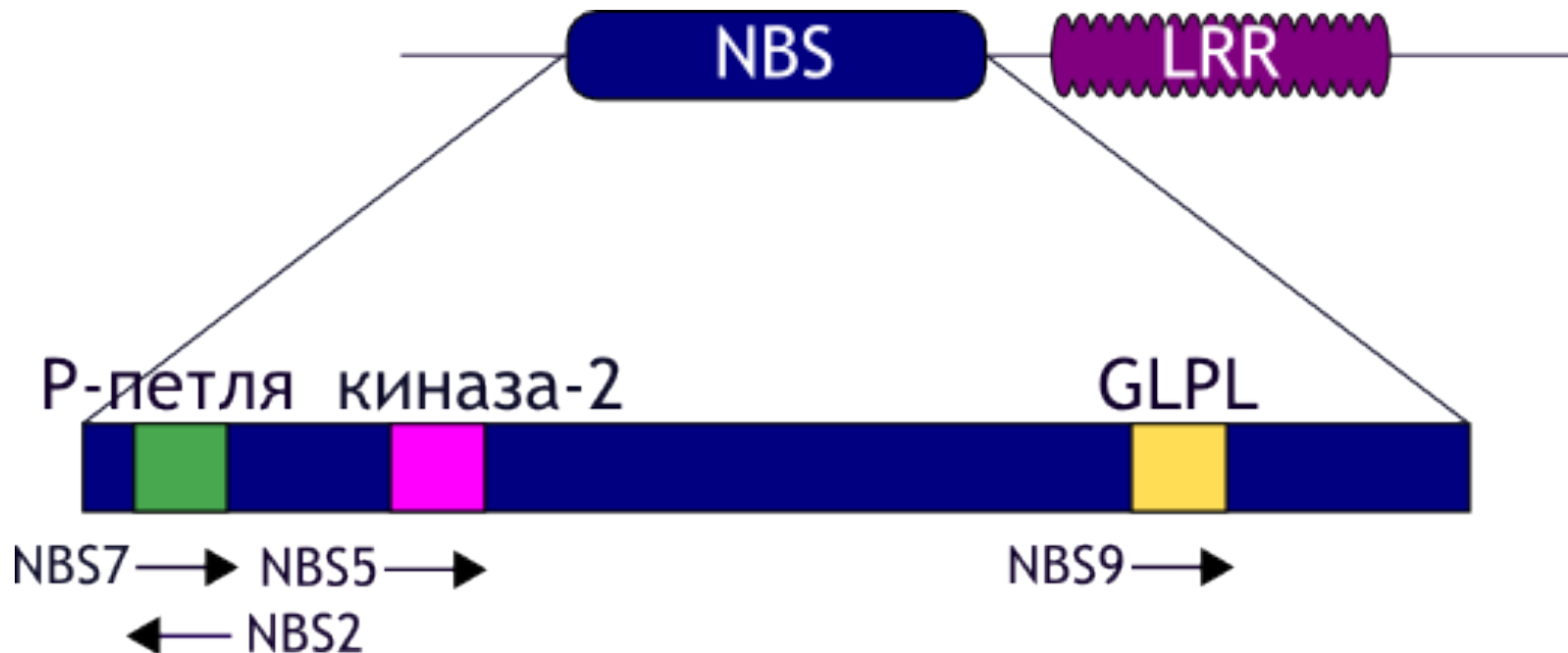
Белки с трансмембранным участком и суперспирализованным (coiled-coil) доменом

5. NBS-LRR

Белки с нуклеотид-связывающим сайтом и доменом с обогащенными лейцином повторами

Исходные данные

Ампликоны NBS-LRR генов **176** образцов растений, принадлежащих к **23** видам рода *Triticum* и **10** видам рода *Aegilops*, полученные методом NBS-профайлинга и секвенированные на Illumina HiSeq, длина парных чтений 100 + 100 нуклеотидов.



Цель работы

Целью данной работы являлось изучение полиморфизма генов устойчивости класса NBS-LRR у видов родов *Triticum* и *Aegilops* с помощью данных высокопроизводительного секвенирования.

Всё начиналось неплохо, но ...

Структура генома видов пшеницы

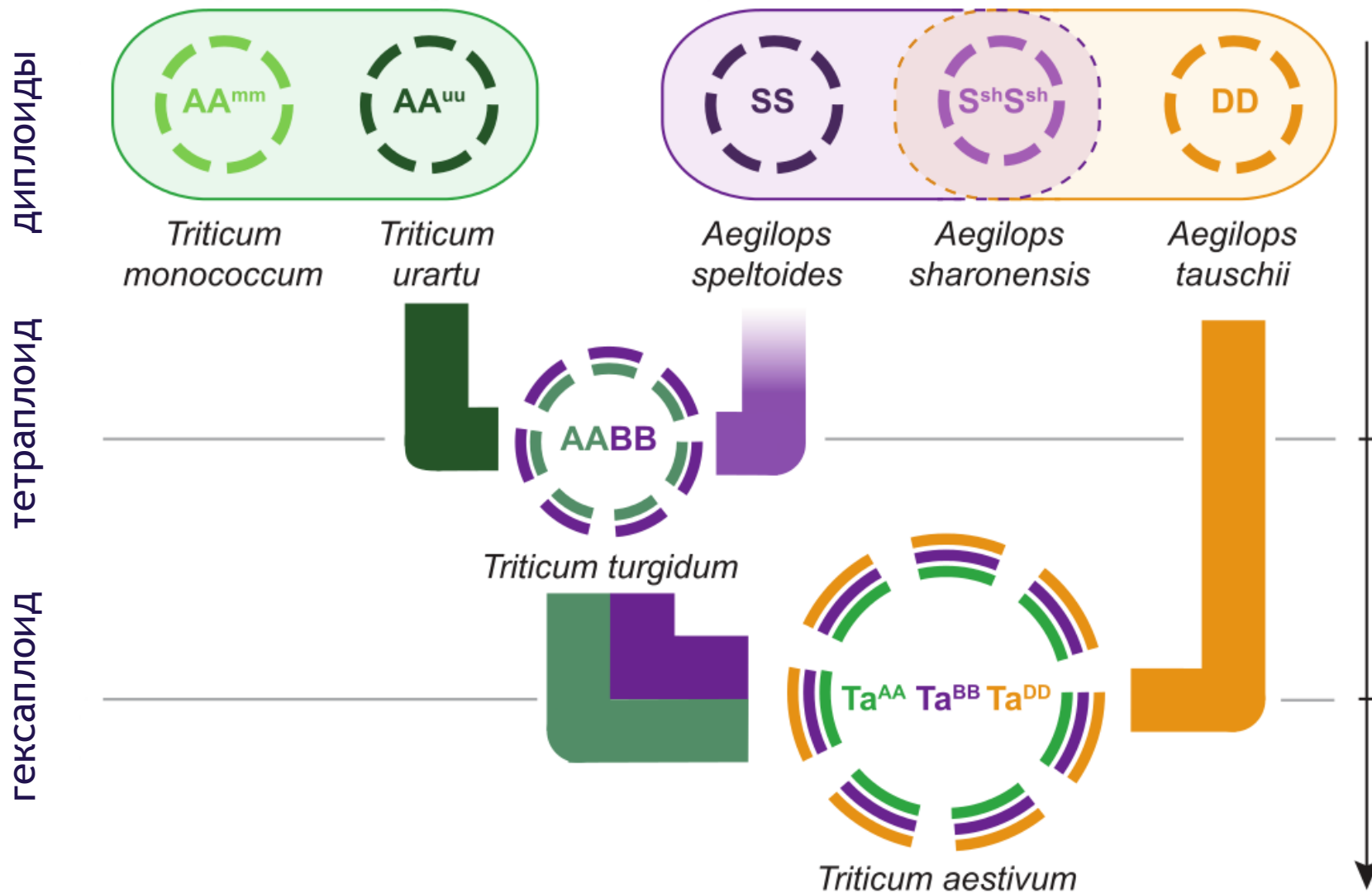


Иллюстрация из статьи "A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome", IWGSC, Science, 2014

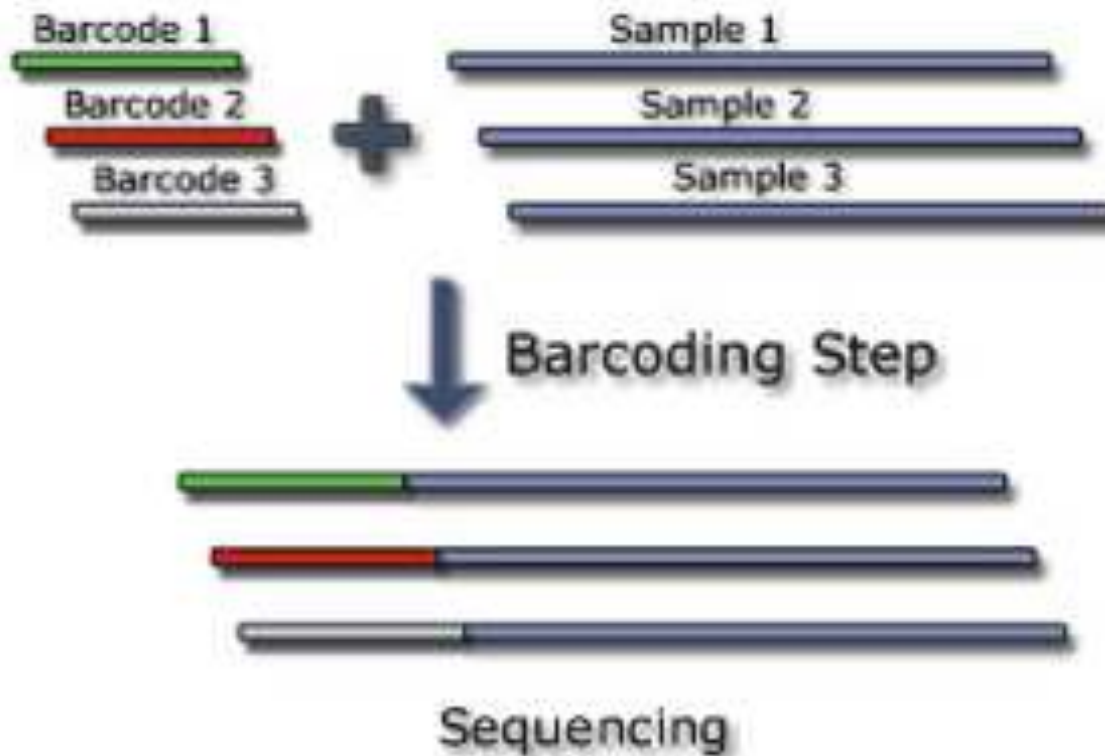
Секвенирование генома *Triticum aestivum*. Всё плохо

Название статьи	Год	% генома	N50, нп	Аннотация в открытом доступе
“Analysis of the bread wheat genome using whole-genome shotgun sequencing”, Brenchley <i>et al.</i> , Nature	2012	2	-	есть
“A chromosome-based draft sequence of the hexaploid bread wheat (<i>Triticum aestivum</i>) genome”, IWGSC, Science	2014	61	2 300	есть (3% всего генома)
“A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome”, Chapman <i>et al.</i> , Genome Biology	2015	56	21 200	нет

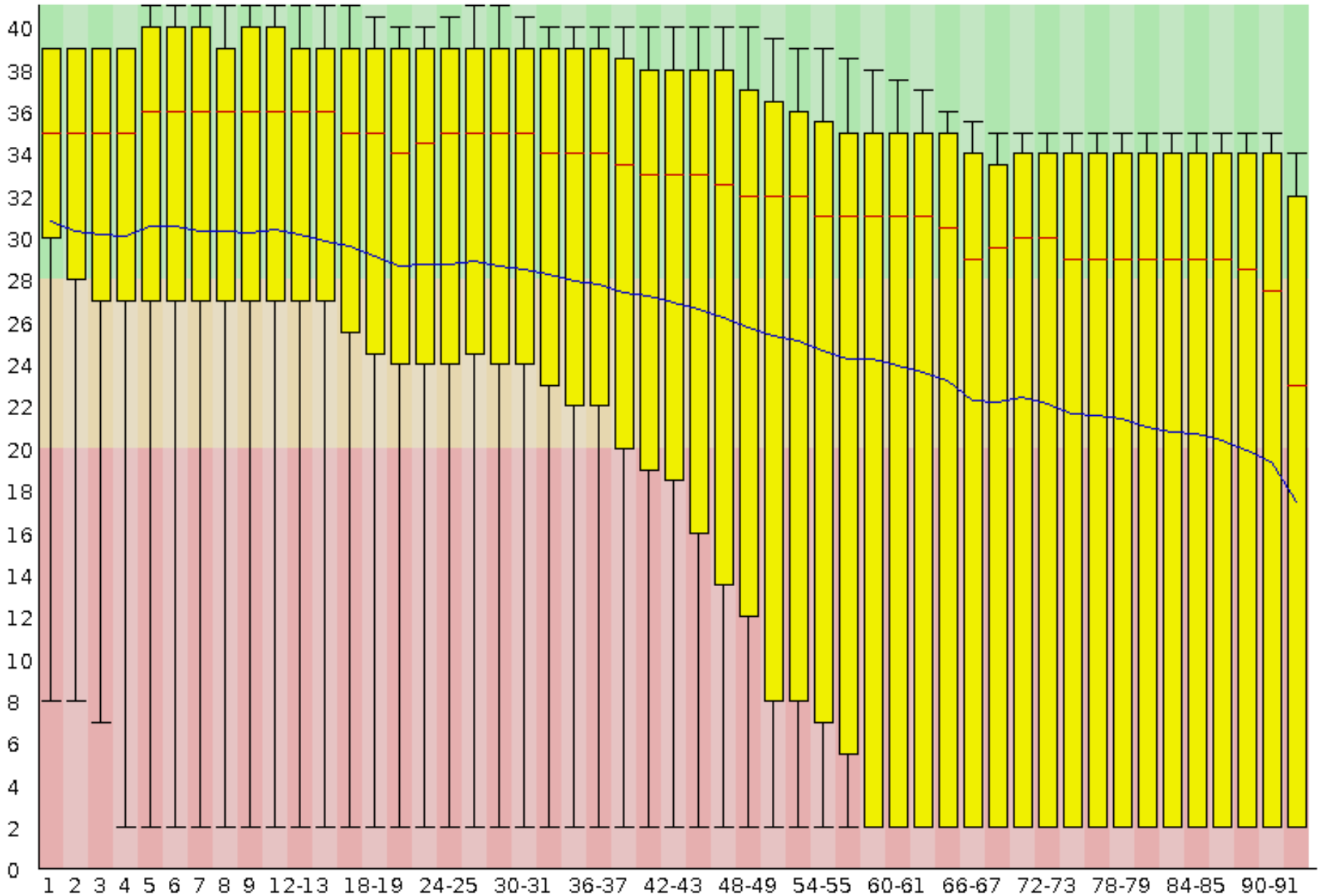
Pipeline для обработки данных



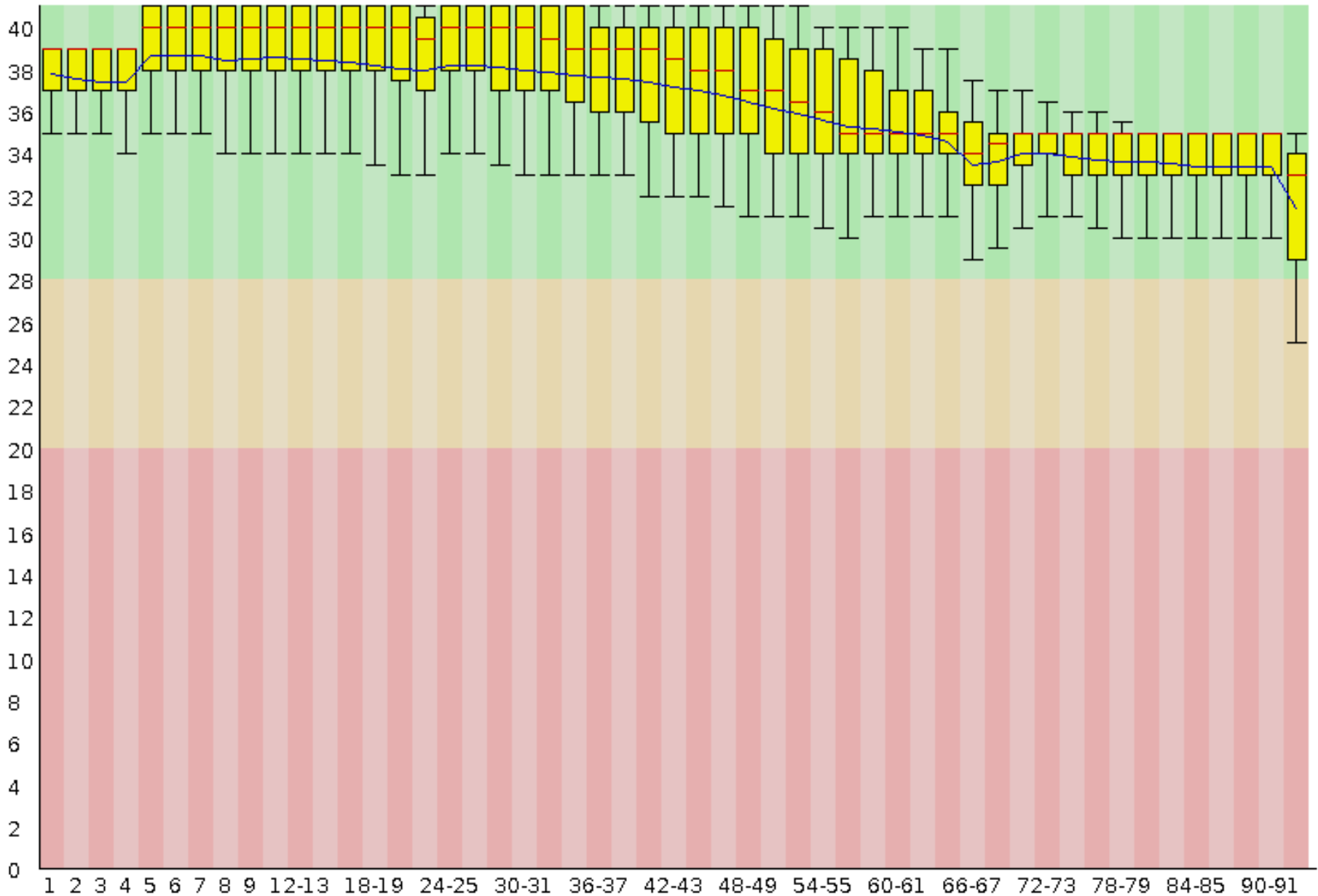
Сортировка рядов по баркодам



Качество ридов до очистки



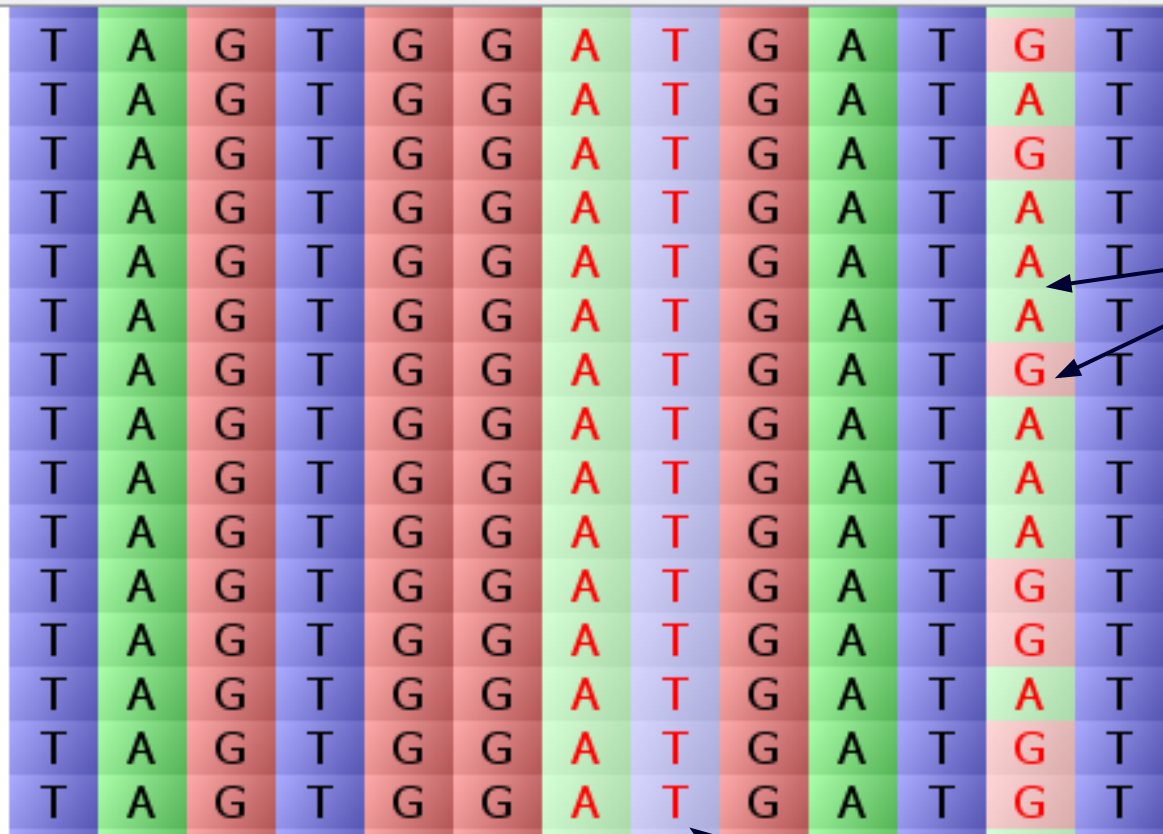
Качество ридов после очистки



Выявление нуклеотидных замен

референсная последовательность

G A T A G T G G T C G A T T T



2 аллеля

картированные чтения

Визуализация выполнена с помощью программы Tablet

полиморфные сайты

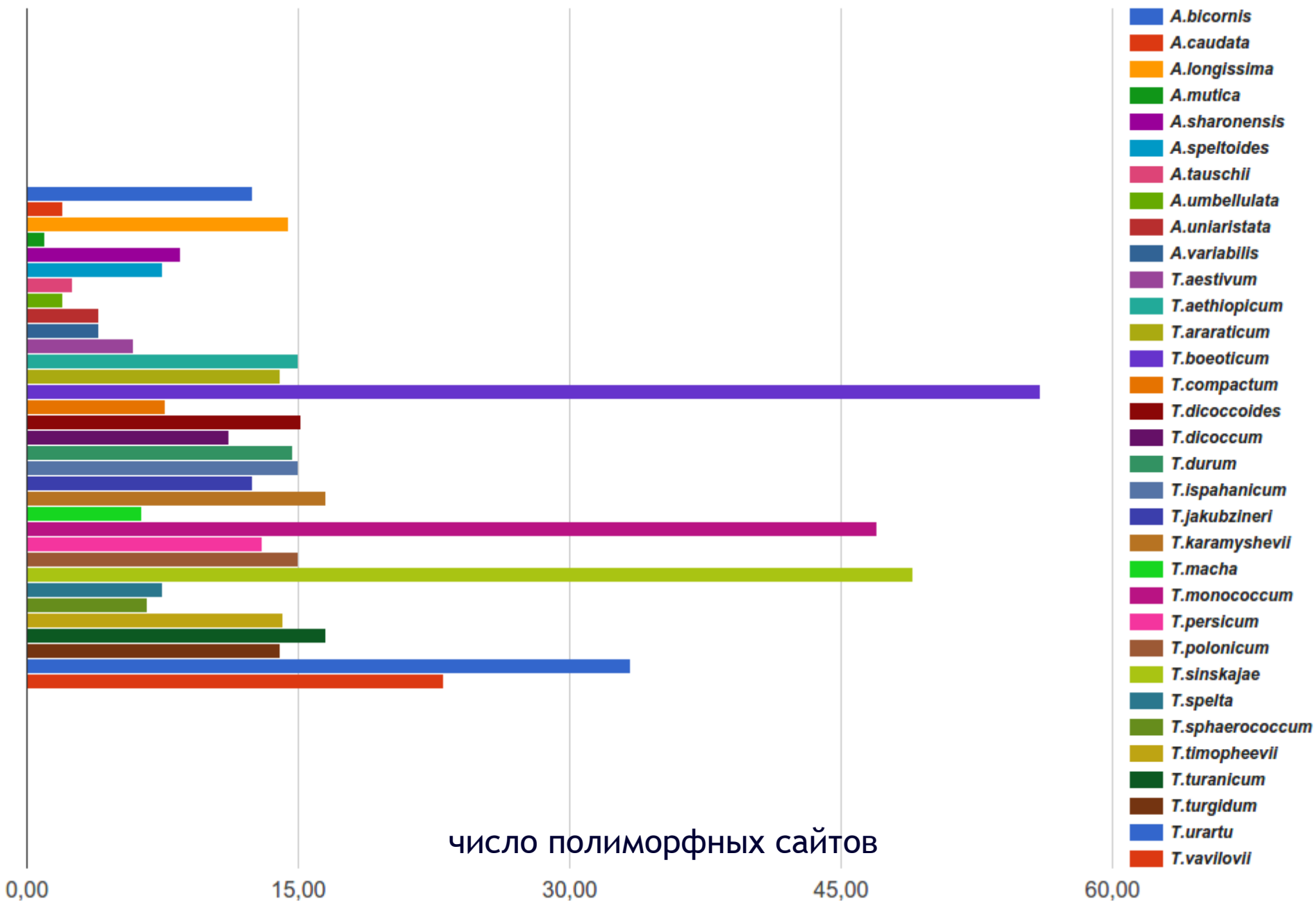
Результаты картирования чтений. Что-то здесь не то...

Референсная последовательность	Размер, Мб	Средняя доля картированных чтений по образцам, %
Аннотированный геном <i>T. aestivum</i> (IWGSC, 2014)	600	64.30
Транскрипты <i>T. aestivum</i> (IWGSC, 2014)	400	7.89
Геном <i>T. aestivum</i> (Chapman <i>et al.</i> , 2015)	9 000	94.89
Последовательности генов <i>NBS-LRR</i> , выделенные на основе гомологии из сборки <i>T. aestivum</i> 2015 года	15	9.00

Результаты анализа полиморфных сайтов

- Было выявлено 446 полиморфных сайтов в 120 генах, содержащих 525 вариантов замен
- Из них 208 - уникальные варианты, встречающиеся только в одном из образцов
- Число замен в образцах варьируется от 0 (*T. aestivum*) до 69 (*T. vavilovi*)
- Среднее число замен на один образец - 21.15

Уровень полиморфизма у различных видов *Aegilops* и *Triticum*?



Уровень полиморфизма у *T. aestivum* и других видов гексаплоидных пшениц

число образцов

30

25

20

15

10

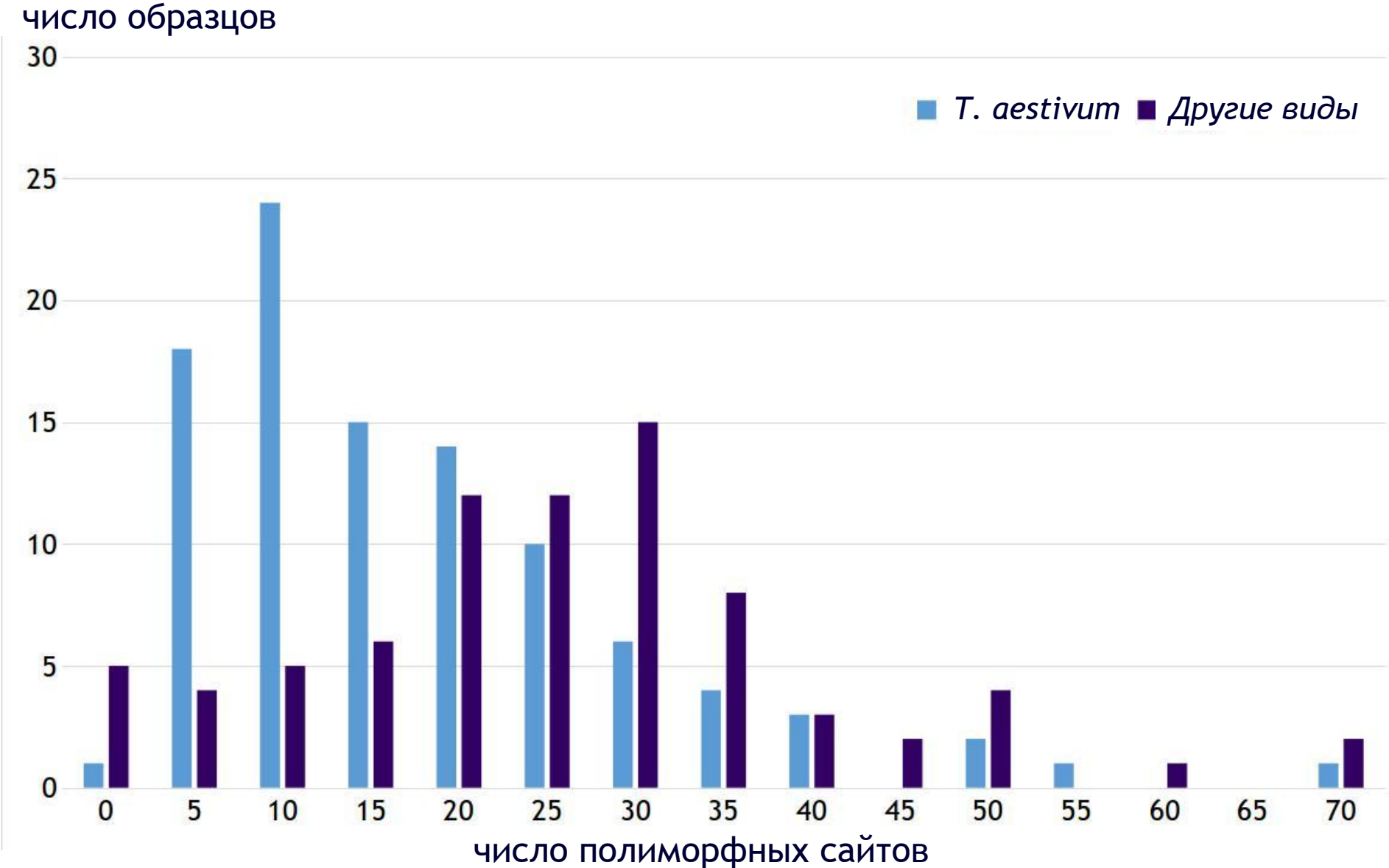
5

0

■ *T. aestivum* ■ Другие виды

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70

число полиморфных сайтов



Выводы

- Разработана и оптимизирована система обработки данных NBS-профайлинга, полученных с использованием высокопроизводительного параллельного секвенирования, для работы с данными низкого качества.
- Показано, что менее 10% полученных методом NBS-профайлинга последовательностей соответствуют целевым NBS-LRR-генам, что ставит под сомнение эффективность использования данного метода для объектов с большими геномами с низким качеством сборки.
- Число полиморфных сайтов в генах NBS-LRR у сортов *Triticum aestivum* находится в пределах от 0 до 68 по сравнению с референсным геномом.

Выводы

- Не стоит связываться с видами с большими и полиплоидными геномами
- Не стоит связываться с видами, чей геном плохо аннотирован
- Нужно хорошо понимать, зачем мы делаем NGS, и какие данные мы туда отправляем
- Перед тем, как отправлять пробы на NGS, надо проверить, что мы секвенируем именно то, что планировали
- Trash in - trash out
- Но, даже если всё идёт очень плохо, надо не отчаиваться, а попросить помощи коллег



I want good data