

Анализ геномов дрожжей из Петергофской генетической коллекции

Антон Матиив

Руководитель:
Юрий Барбитов
Институт биоинформатики
Кафедра генетики и биотехнологии СПбГУ

Nanopore DNA sequencing



Задачи:

- 1) “Полировка” собранного генома из длинных ридов (Nanopolish, Racon)
- 2) Функциональная аннотация на основании референсного штамма
- 3) Использование полученного “полированного” генома в качестве референса для поиска мутаций в других штаммах



Термин "полировка"
применяется в случае
коррекции при помощи
других данных - либо
"сырого сигнала"
(**Nanopolish**), либо
коротких прочтений
(**Racon**).

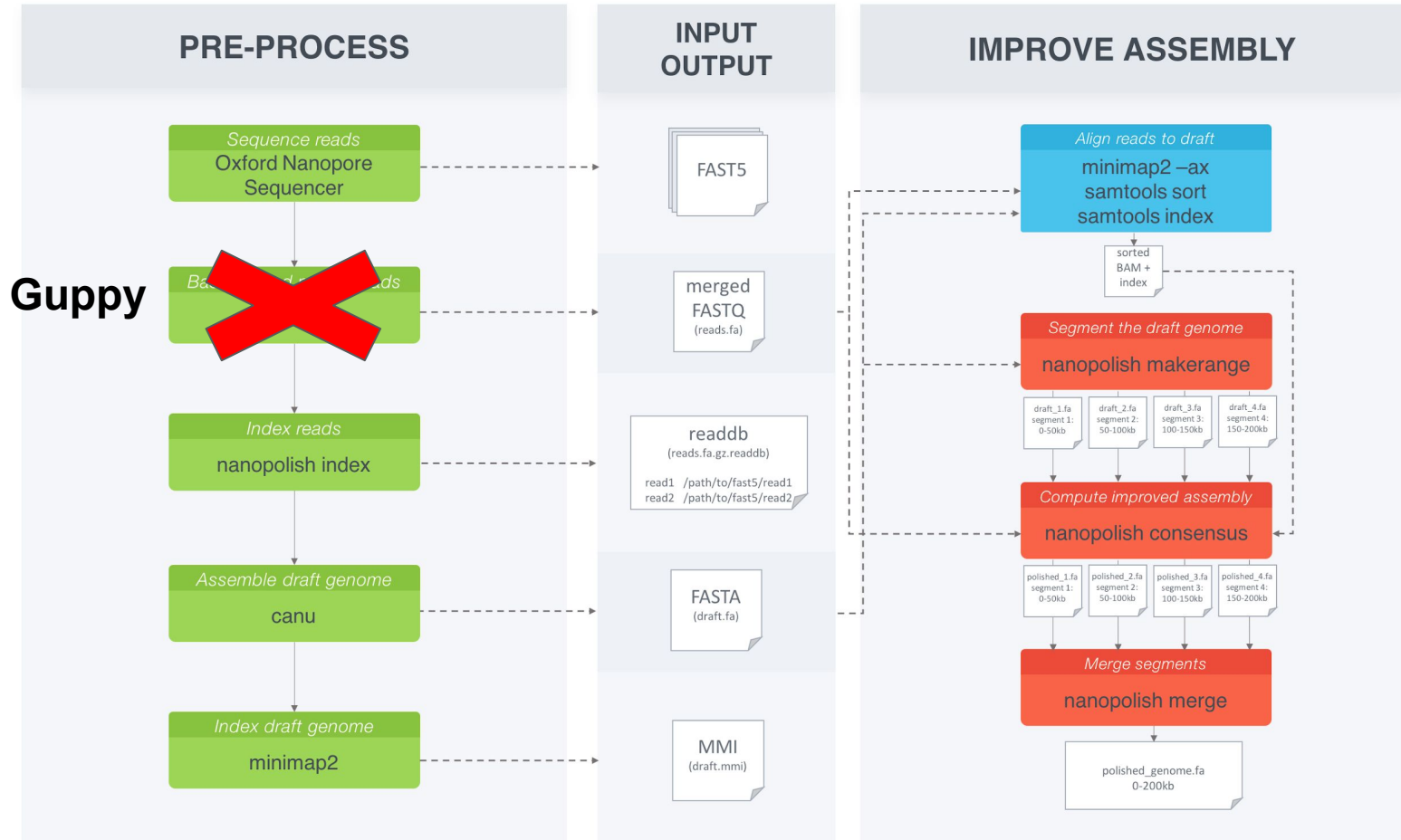


Draft genome assembly



Nanopolish + Racon

Nanopolish pipeline



Оценка сборок с помощью QUAST

	yeast_1d.guppy_213. canu (Draft)	polished_genome	polished_genome _racon
Total length	12396504	12445256	12444850
Largest contig	1095012	1099417	1099192
N50	822302	825460	825513
mismatches per 100 kbp	212.37	207.85	205.12
indels per 100 kbp	343.76	73.09	25.04

Аннотация сборки с помощью Exonerate

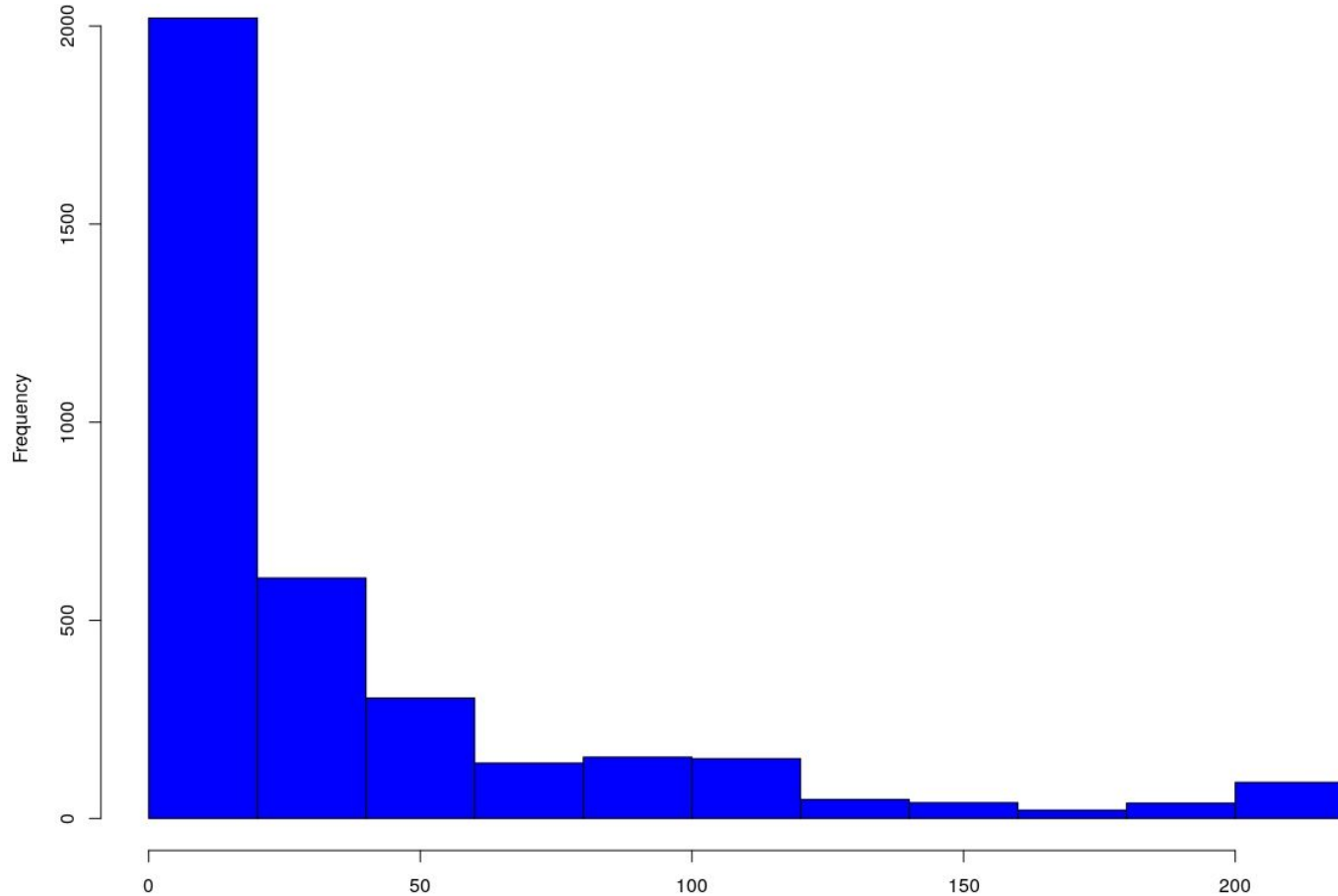
Использовали список аминокислотных последовательностей белков S288С из базы UniProt. Построили базу данных с помощью SnpEff

In progress:

поиск мутаций в образцах с использованием собранного генома в качестве референса



Распределение числа аллелей



Спасибо за внимание!