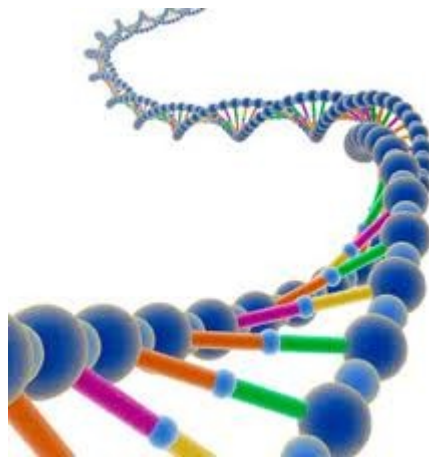


# Стажировки в лаборатории алгоритмической биологии

Проект:  
**QUAST: Quality Assessment Tool**



# Оценка качества

## **Качество сборки:**

набор численных метрик и графиков

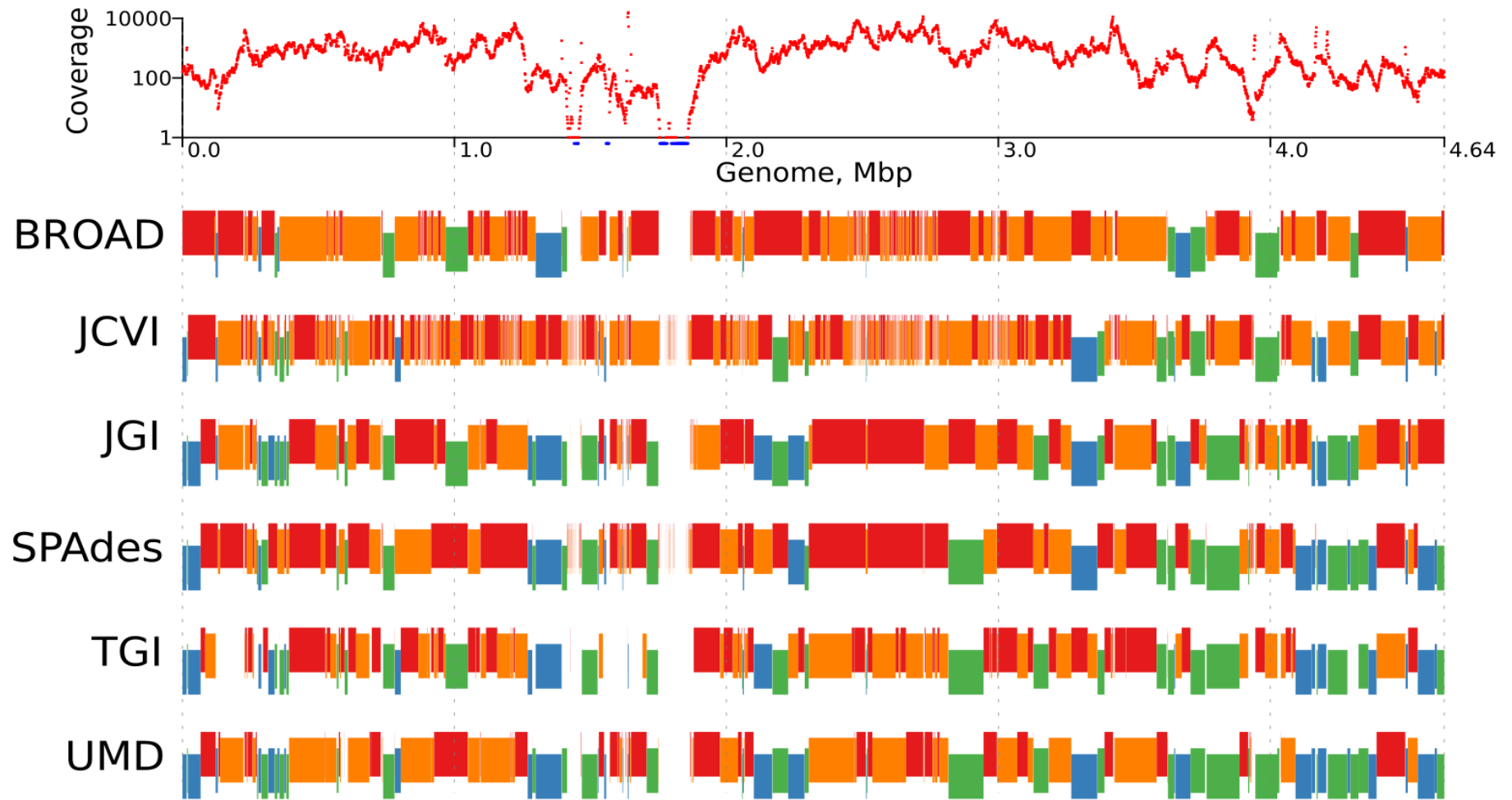
## **Актуальность:**

- Разработка и отладка ассемблера
- Оценка “надежности” ассемблера
- Сравнение ассемблеров между собой

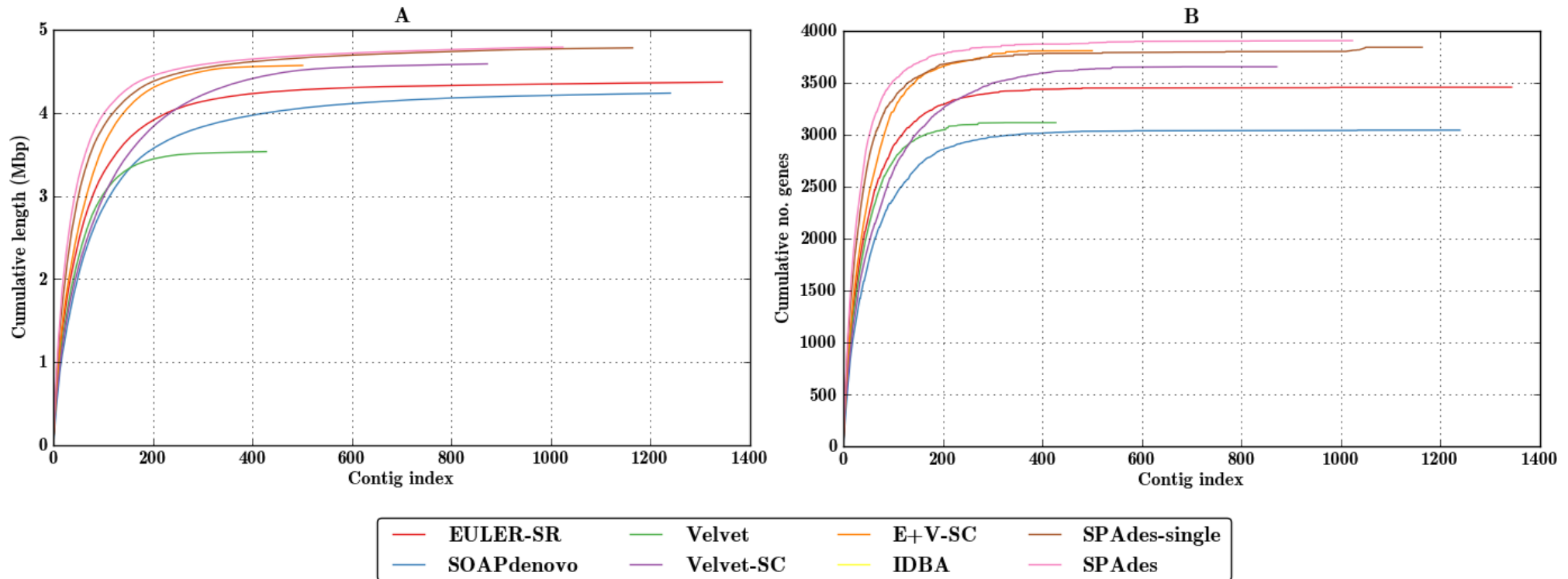
# Что есть

- Консольное приложение на Python с элементами Perl, Java, R
- На входе: сборки, референсный геном, дополнительная информация (например, гены)
- На выходе: отчеты, таблицы, графики, etc

# Пример: Сравнение контигов



# Пример: Графики длины сборки и количества генов от количества контигов



# Пример: Сводная таблица

Assembler	# contigs	N50 (bp)	Largest (bp)	Total (bp)	Covered (%)	Misassemblies	Mismatches (per 100 kbp)	Complete genes
<b>Single-cell <i>E. coli</i> (ECOLI-SC)</b>								
EULER-SR	1344	26662	126616	4369634	87.8	21	11.0	3457
SOAPdenovo	1240	18468	87533	4237595	82.5	13	99.5	3059
Velvet	<b>428</b>	22648	132865	3533351	75.8	2	<b>1.9</b>	3117
Velvet-SC	872	19791	121367	4589603	93.8	2	<b>1.9</b>	3654
E+V-SC	501	32051	132865	4570583	93.8	2	6.7	3809
SPAdes-single reads	1164	42492	166117	4781576	<b>96.1</b>	<b>1</b>	6.2	3888
SPAdes	1024	<b>49623</b>	<b>177944</b>	4790509	<b>96.1</b>	<b>1</b>	5.2	<b>3911</b>
<b>Normal multicell sample of <i>E. coli</i> (ECOLI-MC)</b>								
EULER-SR	295	<b>110153</b>	221409	4598020	99.5	10	5.2	4232
IDBA	<b>191</b>	50818	164392	4566786	99.5	4	1.0	4201
SOAPdenovo	192	62512	172567	4529677	97.7	1	26.1	4141
Velvet	198	78602	196677	4570131	<b>99.9</b>	4	1.2	4223
Velvet-SC	350	52522	166115	4571760	<b>99.9</b>	<b>0</b>	1.3	4165
E+V-SC	339	54856	166115	4571406	<b>99.9</b>	<b>0</b>	2.9	4172
SPAdes-single reads	445	59666	166117	4578486	<b>99.9</b>	<b>0</b>	<b>0.7</b>	4246
SPAdes	195	86590	<b>222950</b>	4608505	<b>99.9</b>	2	3.7	<b>4268</b>

# Что нужно

## **Техническая часть:**

- web-service (Django),
- реализация новых метрик (Python),
- etc

## **Идеологическая часть:**

- работа с мультихромосомными геномами,
- оценка качества при отсутствии референсного генома,
- etc

**Спасибо за внимание!**

**Вопросы?**

**<http://bioinf.spbau.ru/int>  
[alexei.gurevich@gmail.com](mailto:alexei.gurevich@gmail.com)**