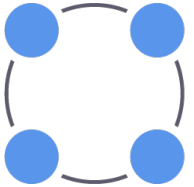


BIOINFORMATICS INSTITUTE

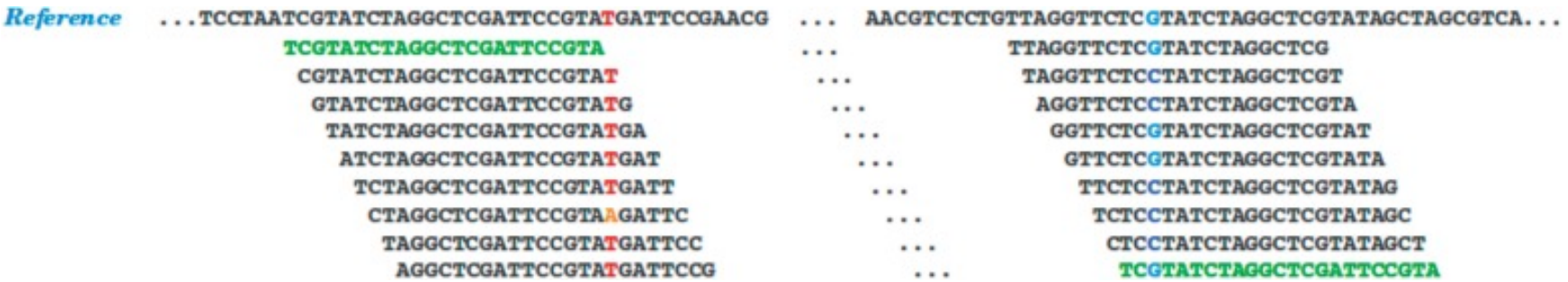
Расширение функциональности инструмента для построения функции выравниваемости генома

Руководитель: Евгений Бакин
(Институт биоинформатики)

Студент: Елизавета Скалон

Genome Mappability Score (GMS)

отражает вероятность того, что рид может быть однозначно выровнен на данную позицию генома



GMS низкий — на данную позицию генома может выровняться много ридов
→ на данную позицию может быть выровнен неправильный рид
→ неверная интерпретация результатов выравнивания

GMS высокий — данная позиция генома уникальна
→ выравниванию можно доверять

Существующие реализации:

Genome Mappability Analyzer

Bioinformatics. 2012 Aug 15; 28(16): 2097–2105.
Published online 2012 Jul 4. doi: 10.1093/bioinformatics/bts330
PMCID: PMC3413383

Genomic dark matter: the reliability of short read mapping illustrated by the genome mappability score

Hayan Lee[1,2,*] and Michael C. Schatz[1,2]

«GMStool»

Hardware-effective Tool for Genome Mappability Score Estimation	Feb 2017	Jun 2017	Евгений Бакин Наталья Зорина	Александр Предеус

Цель:

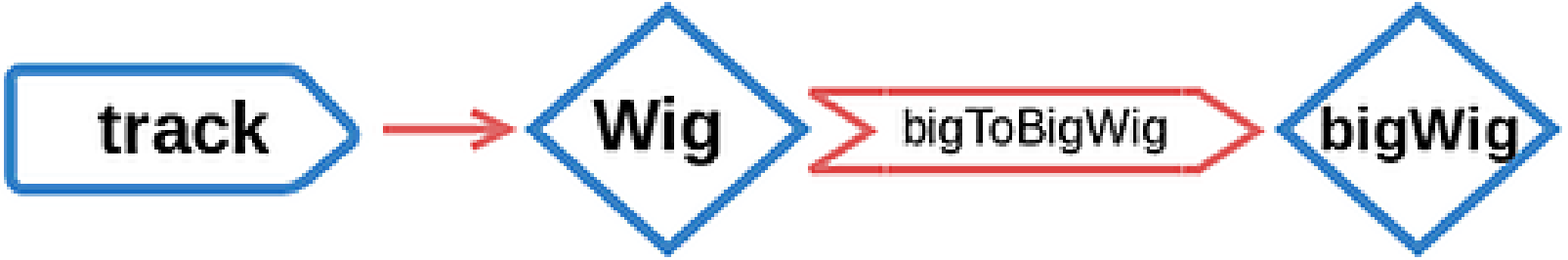
Расширить функциональность тула в части работы с различными форматами данных

Задачи:

- Добавить возможность выводить данные в форматы BED, bigBed и TDF
- Ускорить процесс выгрузки данных
- Реализовать запись данных непосредственно в формат bigWig

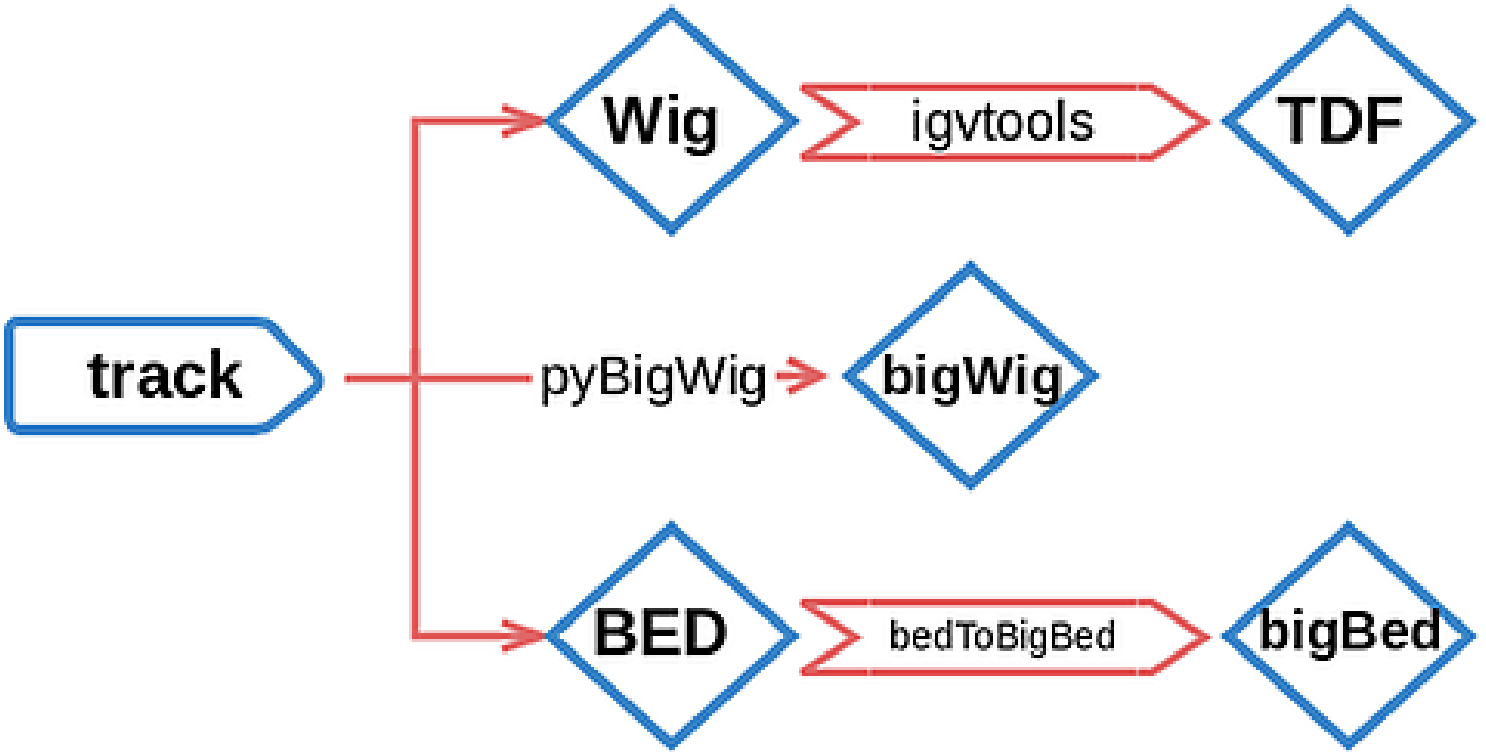
Итоги
Разнообразие форматов

Раньше:
- Wig
- bigWig

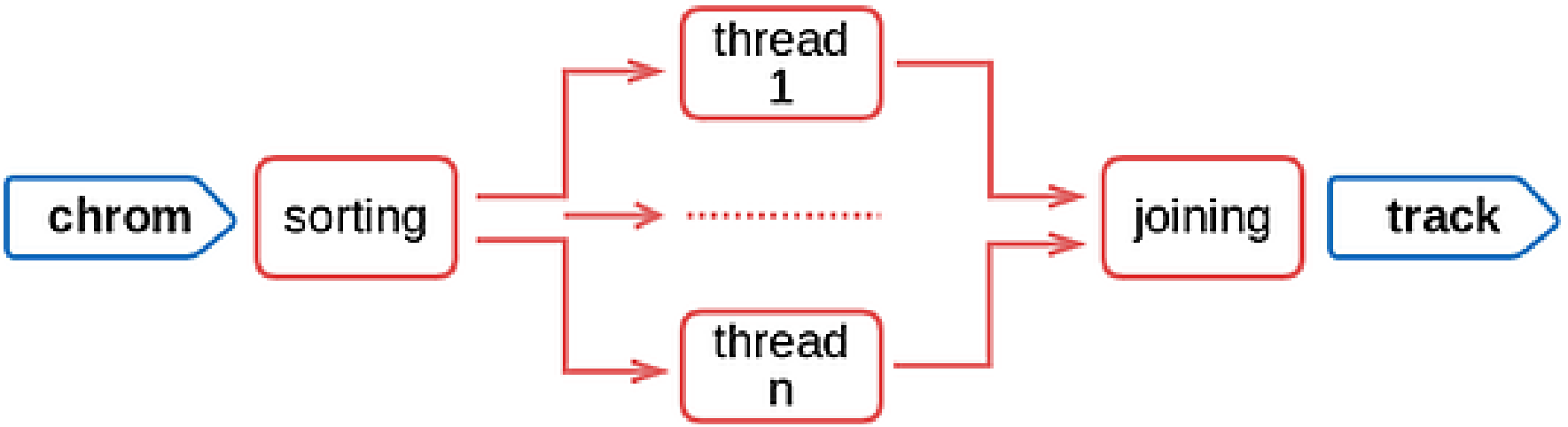


Итоги
Разнообразие форматов

Сейчас:
- Wig
- bigWig
- BED
- bigBed
- TDF



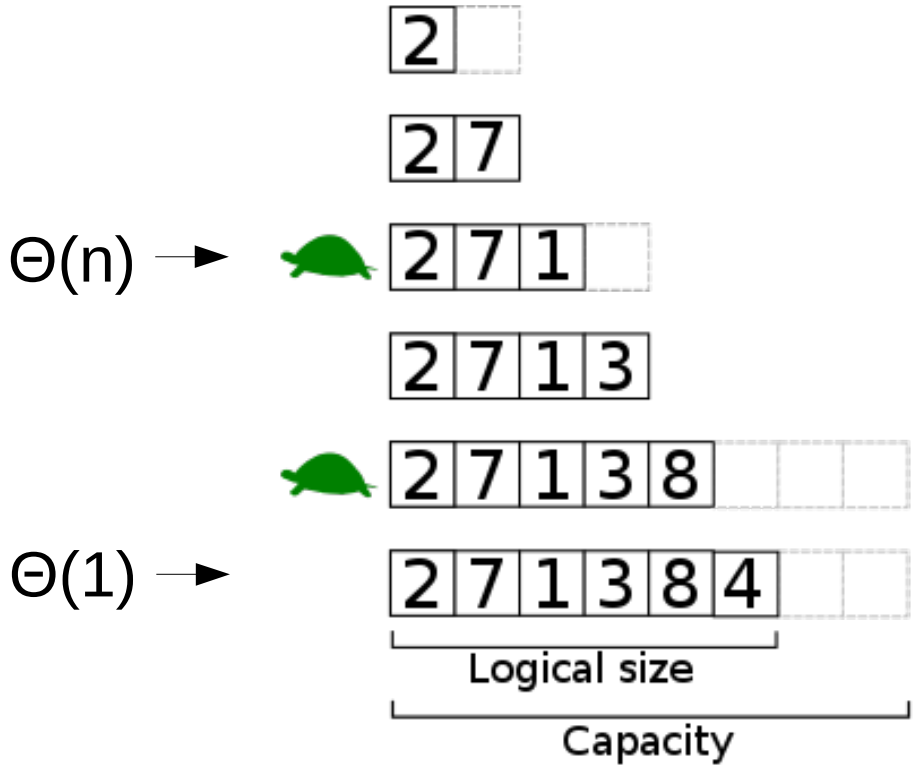
Итоги
Скорость: multiprocessing



* n, то есть количество потоков, задается пользователем.

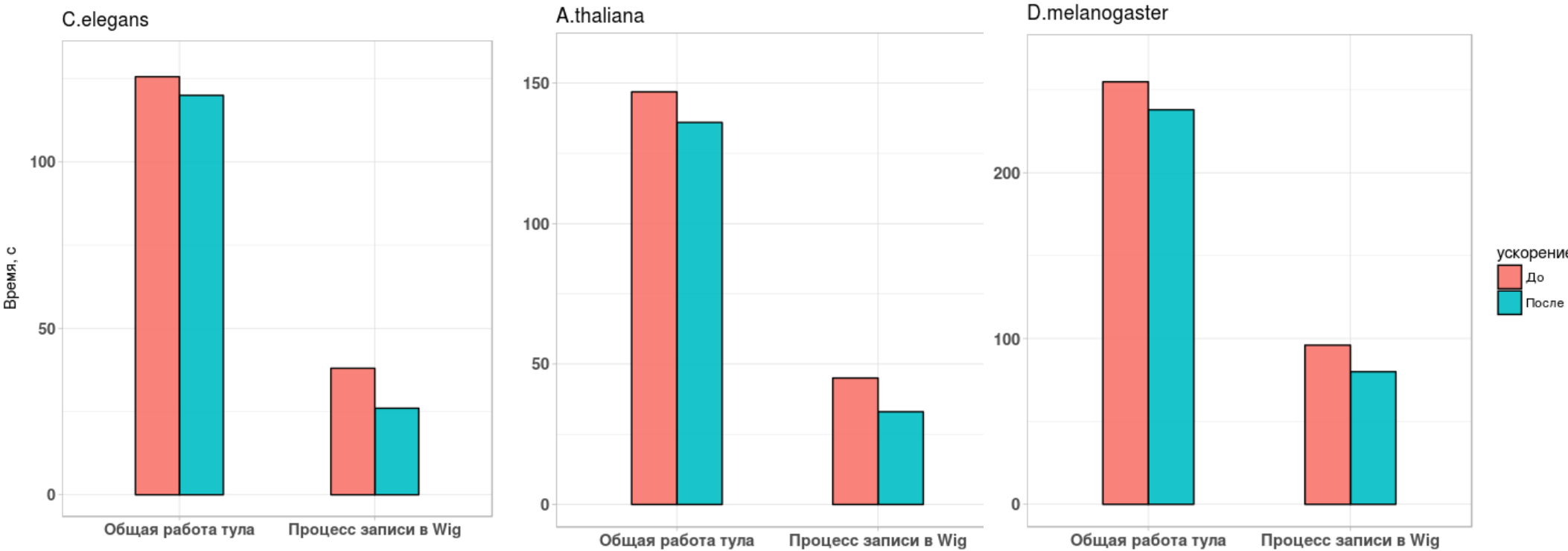
Итоги

Скорость: геометрическое расширение памяти



Итоги

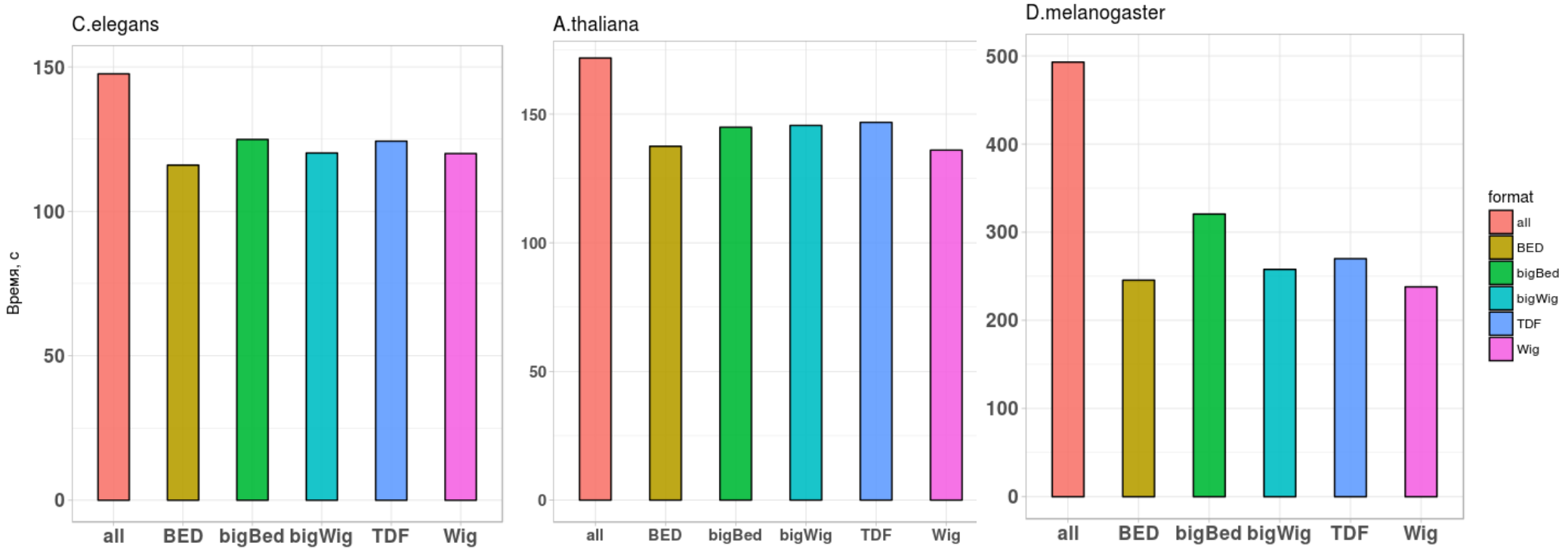
Увеличение скорости на примере экспорта в Wig



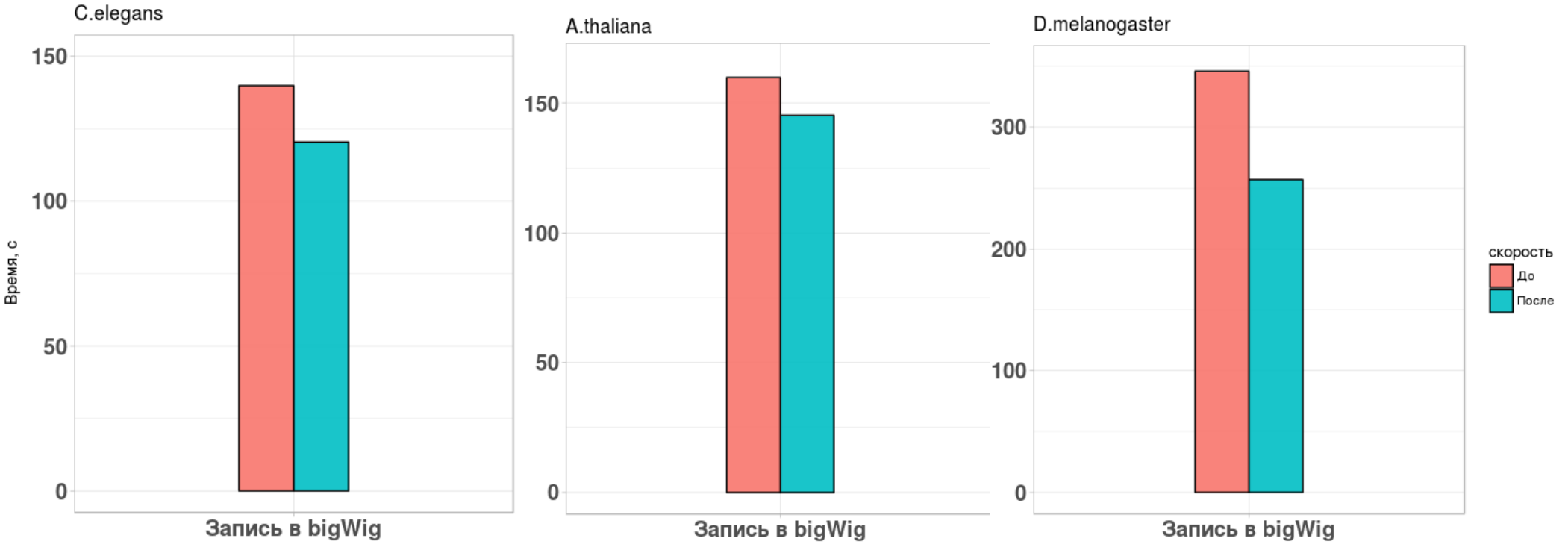
* подсчет проводился на 4-ядерном процессоре

Итоги

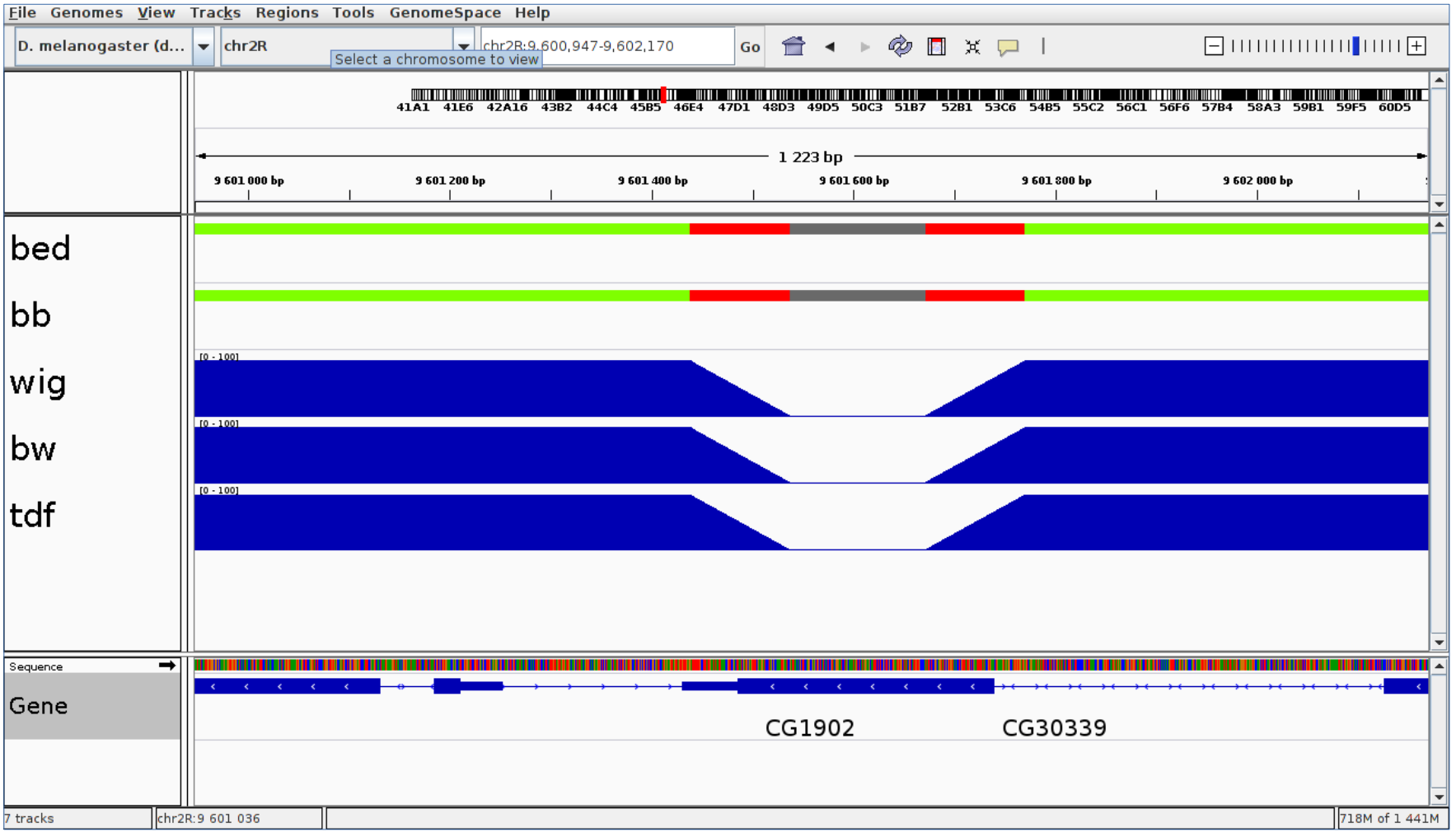
Скорость: формирование единого трека для последующей записи в разные форматы



Итоги
Скорость: экспорт напрямую в bigWig



Пример работы тула



Выводы

Расширен функционал тула в части экспорта полученных результатов:

- Сейчас возможен вывод в 5 различных форматов: Wig, bigWig, BED, bigBED, TDF. Каждый из них имеет свои преимущества для пользователя.
- Скорость выгрузки уменьшена за счет реализации многопоточного режима, а также за счет геометрического расширения памяти.
- Реализован экспорт данных непосредственно в формат bigWig

Планы на будущее:

- Усовершенствовать остальную часть тула (мультипроцессинг и тд.)

Спасибо за внимание!