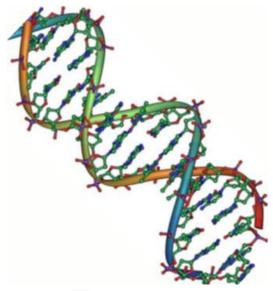


Разработка алгоритма сборки генома для вычислительных систем экзафлопсного уровня производительности



Геном



Чтение генома

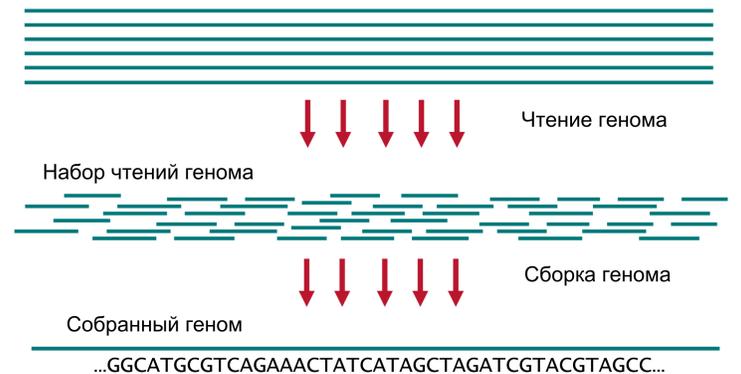


Компьютерная обработка данных

- Геном – последовательность из четырех типов нуклеотидов: А, G, С и Т
- Геномы разных живых существ имеют разный размер – от нескольких тысяч нуклеотидов до сотен миллиардов нуклеотидов
- Геном человека – 3 миллиарда нуклеотидов
- Чтения генома имеют длину примерно 100 нуклеотидов



Paris Japonica – геном в 150 миллиардов нуклеотидов

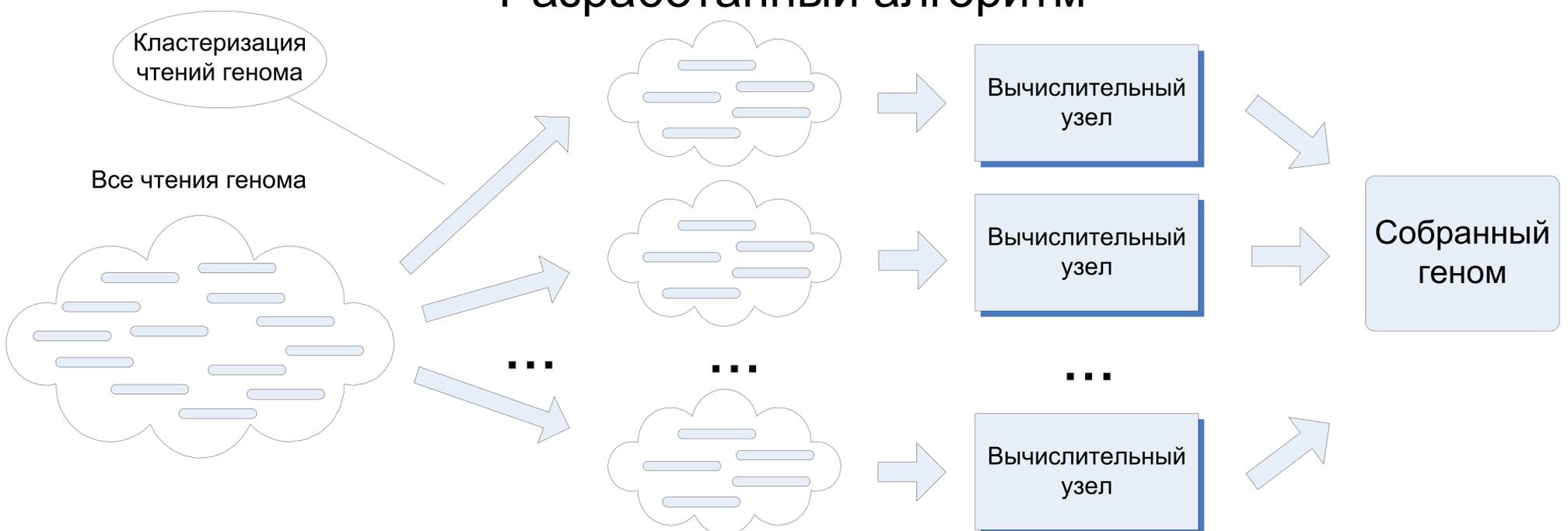


- Производительность устройств для чтения генома увеличивается в 10 раз каждые 18 месяцев
- Производительность компьютеров за тот же срок увеличивается в 2 раза
- Компьютеры экзафлопсной производительности (10^{18} операций в секунду) планируется построить к 2018 году

	Суперкомпьютер сегодня	Суперкомпьютер 2018 года
Пиковая производительность	2 петафлопса	1 экзафлопс
Объем памяти	0.3 петабайта	32-64 петабайта
Производительность вычислительного узла	125 гигафлопс	1-10 терафлопс
Число ядер на одном узле	12	1000 – 10000
Число вычислительных узлов	18700	100000 – 1000000
Объем хранилища данных	15 петабайт	500-1000 петабайт

- Экзафлопсные суперкомпьютеры будут существенно отличаться от сегодняшних
- Вывод – необходимы новые алгоритмы обработки геномных данных

Разработанный алгоритм



Результаты

- Разработан алгоритм сборки генома для вычислительной системы экзафлопсного уровня производительности
- Теоретическая оценка ускорения показывает, что при переходе от 30 вычислительных узлов к 1 млн. узлов скорость обработки данных повышается в 1000 раз, а для отдельных этапов алгоритма – в 16000 раз

Государственный контракт №07.514.11.4010

Лаборатория «Алгоритмы сборки геномных последовательностей»
НИУ ИТМО

<http://genome.ifmo.ru>
genome@mail.ifmo.ru

