

Практикум 15

Работаем в папке /mnt/scratch/NGS/alexrom/pr15
Все адаптеры я записала в файл adapters.fa

1. Команда: TrimmomaticSE -threads 15 -phred33 SRR4240360.fastq.gz
SRR4240360_trimmed.fastq ILLUMINACLIP:adapters.fa:2:7:7

На выходе: 8254632 Surviving: 8212774 (99.49%) Dropped: 41858 (0.51%)

Как можно заметить, 0.51% последовательностей чтений оказались остатками адаптеров.

Далее надо удалить с правых концов чтений нуклеотиды с качеством ниже 20 и нуклеотиды с длиной меньше 32 нуклеотидов

Команда: TrimmomaticSE -threads 15 -phred33 SRR4240360_trimmed.fastq
SRR4240360_final.fastq TRAILING:20 MINLEN:32

На выходе: 8212774 Surviving: 7915474 (96.38%) Dropped: 297300 (3.62%)

Было удалено 297300 чтений (3.62%).
Файл до финальной очистки: 828M
Файл после финальной очистки: 797M

2. Программа velveth

Команда: velveth results 31 -short -fastq SRR4240360_final.fastq

На файла SRR4240360_final.fastq команда подготавливает k-меры длины k=31
results - папка с результатами
31 - длина k-мера
-short - рассматриваются только короткие и не парные чтения.

Вывод программы: 7915474 sequences in total.

3. Программа velvetg

Сборка на основе созданных k-меров

Команда: velvetg results

results - папка с результатом

На выходе: Final graph has 603 nodes and n50 of 43070, max 113474, total 678075, using 0/7915474 reads

N50 = 43070

Ищем самые длинные контиги. Для это открываем файл stats.txt

ID	Length	Cov (покрытие)
1	113474	33.525460
4	64155	35.847323
5	83603	33.646065

Контиги с аномально большим покрытием

ID	Cov	Length
173	134953	1
345	297.5	2
394	590	1

Контиги с аномально маленьким покрытием

ID	Cov	Length
520	1.75	8
549	1.83	6
565	1.61	3

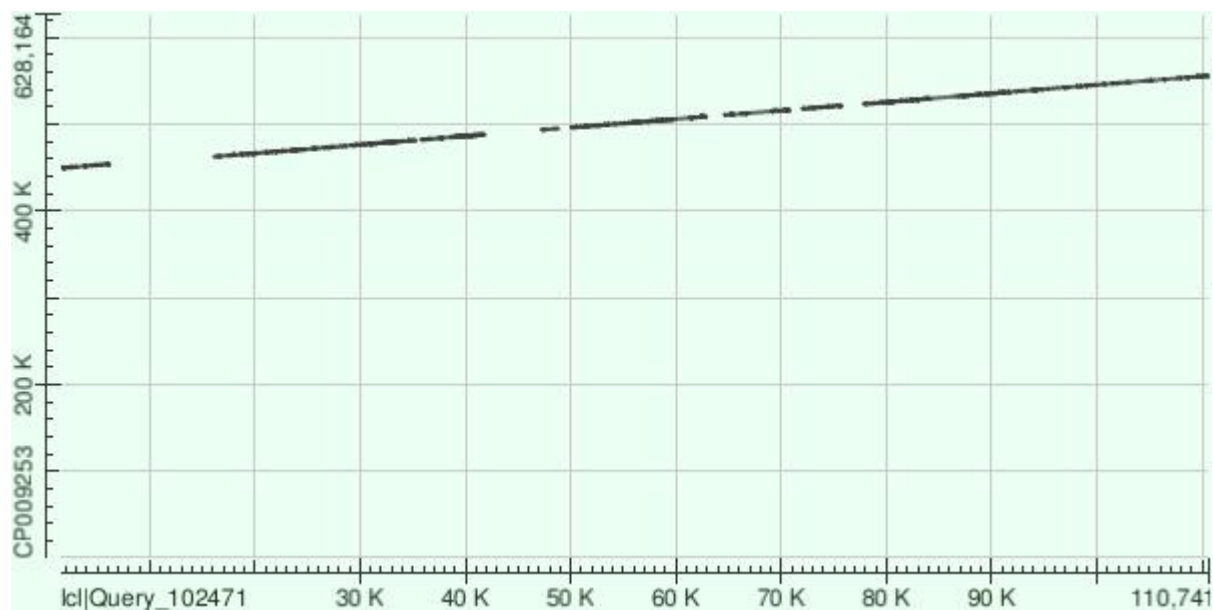
4. Анализ BLAST

Программой megablast я сравнила каждый из трёх самых длинных контигов (ID: 1, 4, 5) с хромосомой *Buchnera aphidicola* (GenBank/ EMBL AC — CP009253).

Вот, что получилось

1) Для континга 1

Участок	Идентичные участки	Гэпы
528794 to 550219	17688/21721(81%)	545/21721(2%)
550361 to 555905	4573/5655(81%)	127/5655(2%)
467412 to 474667	5691/7389(77%)	208/7389(2%)
500370 to 508806	6513/8614(76%)	345/8614(4%)
510438 to 516539	4894/6238(78%)	194/6238(3%)
523105 to 528679	4373/5687(77%)	210/5687(3%)
462496 to 467421	3862/5015(77%)	162/5015(3%)
481997 to 488106	4621/6238(74%)	308/6238(4%)
474844 to 480660	4426/5971(74%)	250/5971(4%)
449411 to 454069	3571/4732(75%)	152/4732(3%)
517766 to 521500	2922/3782(77%)	99/3782(2%)
496111 to 500325	3253/4323(75%)	153/4323(3%)
493487 to 494864	1108/1384(80%)	13/1384(0%)
480874 to 481545	564/686(82%)	20/686(2%)
495033 to 495148	107/120(89%)	5/120(4%)



б) для контига 4

Участок	Идентичные участки	Гэпы
2004 to 11103	7229/9223(78%)	256/9223(2%)
613658 to 620926	5845/7379(79%)	184/7379(2%)
599832 to 604795	3946/5046(78%)	170/5046(3%)
621055 to 627104	4678/6173(76%)	248/6173(4%)
23067 to 28363	4159/5433(77%)	219/5433(4%)
17962 to 20182	1902/2231(85%)	30/2231(1%)
4727 to 17919	2451/3226(76%)	88/3226(2%)
30013 to 32745	2150/2777(77%)	84/2777(3%)
20358 to 22183	1509/1851(82%)	51/1851(2%)
611633 to 613671	1625/2086(78%)	66/2086(3%)
13994 to 14465	393/478(82%)	9/478(1%)
611229 to 611524	236/297(79%)	2/297(0%)



в) для контига 5

Участок	Идентичные участки	Гэпы
127825 to 140555	9751/13010(75%)	548/13010(4%)
153752 to 161738	6355/8168(78%)	264/8168(3%)
144368 to 151796	5859/7536(78%)	243/7536(3%)
101712 to 108876	5567/7274(77%)	215/7274(2%)
161898 to 166752	3911/4914(80%)	112/4914(2%)
166750 to 173180	4967/6517(76%)	159/6517(2%)
126623 to 127815	1004/1199(84%)	11/1199(0%)
98408 to 9930	731/901(81%)	9/901(0%)

