Processing RNA-seq Data

2020-07-23

Getting Files Off the Server

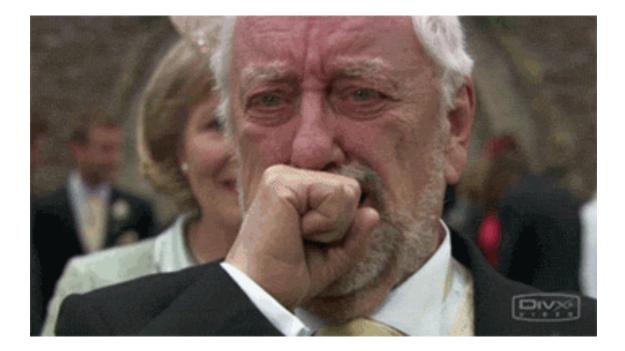
Windows Mac . WinSCP Free FTP CLIENT

Cyberduck

Setting Up a Project

1.For your future self

1.For your future self



1.For your future self



For your future self
 For everybody

else



For your future self
 For everybody else
 For your next step
 in your career



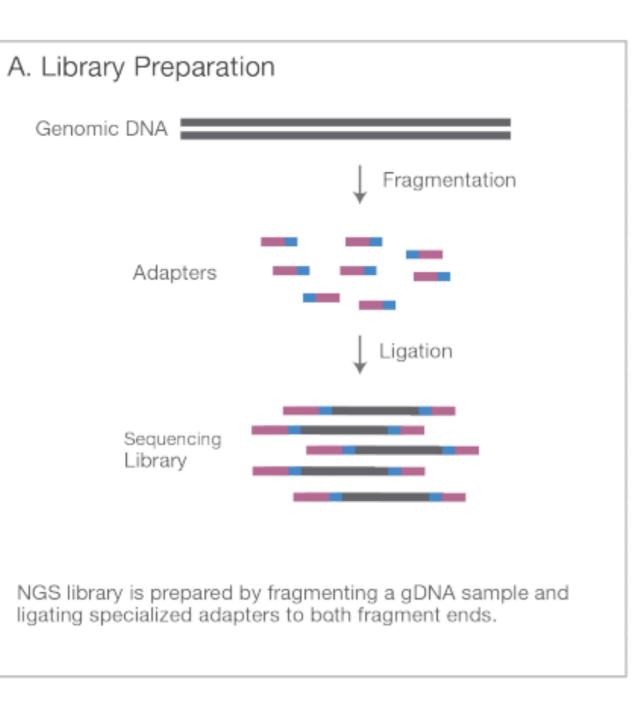
Setting Up Project Documentation

- 1. Create a new folder
- 2. Open a plain text file in a new to take notes in / document your work
- 3. Make that folder into a Git repository and back it up to GitHub

Quick Review: How does Illumina sequencing work?

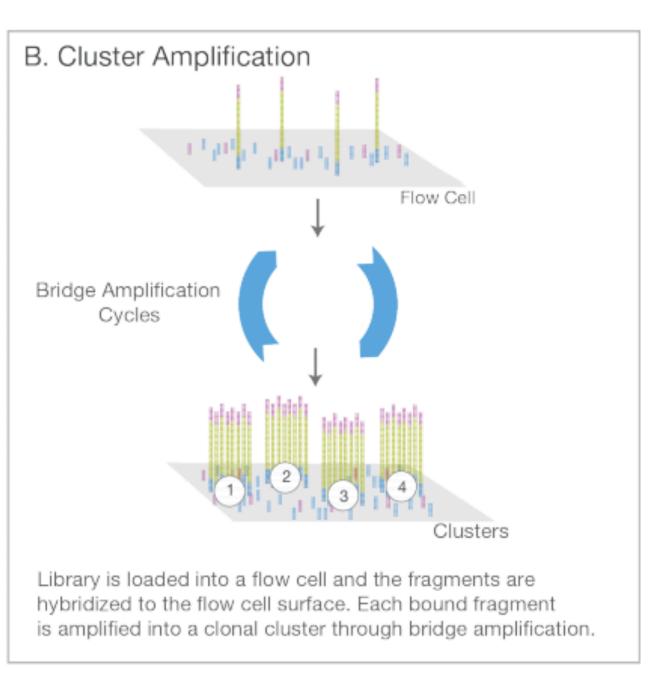
Illumina Sequencing

- General overview
- For RNA-seq, extract RNA and remove ribosomal RNA as well

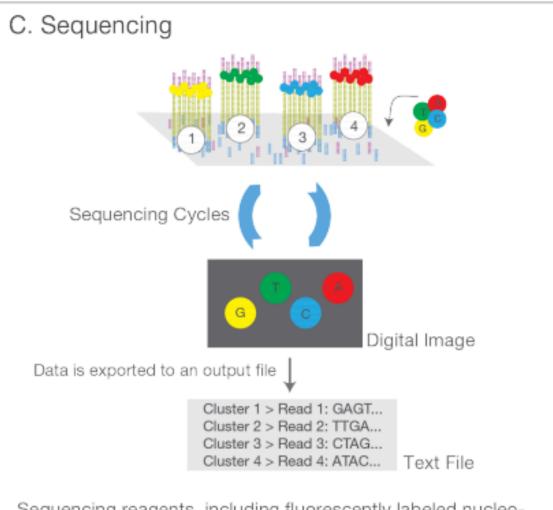


Illumina Sequencing

Make tiles of identical DNA to read



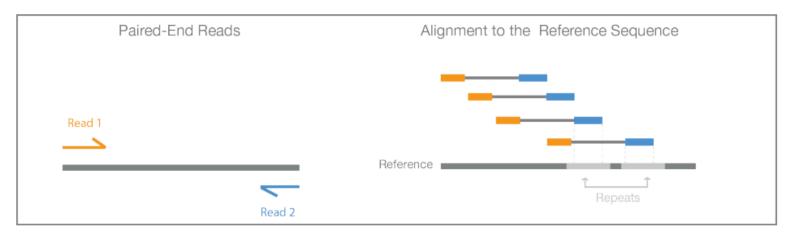
Illumina Sequencing



Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated "n" times to create a read length of "n" bases.

Paired-End Sequencing

- Sequence both ends of the fragment
- Because sequencing is always 5' to 3', the read pairs will be in the opposite orientation
- 90% of the time, the programs you use will be aware of the difference in orientation and take care of it for you
- Because the distance between the pairs is known (depends on the sequence length you asked for) mapping is more accurate, especially in highly repetitive regions of the genome
- For RNA-seq, paired end reads are necessary if you want to look at alternative splicing
- More expensive than single end sequencing



What does raw sequencing data look like?

• Fastq files (usually) end in either fastq.gz or fq.gz (or they can be missing the .gz extension)

- Fastq files (usually) end in either fastq.gz or fq.gz (or they can be missing the .gz extension)
- File names will have some combination of the following information (depends on the sequencer):
 - Sample ID
 - Lane
 - Read number
 - Unique ID from the company or sequencer

- Fastq files (usually) end in either fastq.gz or fq.gz (or they can be missing the .gz extension)
- File names will have some combination of the following information (depends on the sequencer):
 - Sample ID
 - Lane
 - Read number
 - Unique ID from the company or sequencer

[[kkeith@cbix rnaseq_data]\$ ll

```
total 46700
```

```
-rw-r--r--. 1 kkeith research 4276449 Dec 18 10:55 dac1_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 4441834 Dec 18 10:55 dac1_chr21_R2.fastq.gz
-rw-r--r--. 1 kkeith research 4118184 Dec 18 10:55 dac2_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 4296786 Dec 18 10:55 dac2_chr21_R2.fastq.gz
-rw-r--r--. 1 kkeith research 4336091 Dec 18 10:56 dac3_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 4519748 Dec 18 10:56 dac3_chr21_R2.fastq.gz
-rw-r--r--. 1 kkeith research 3652875 Dec 18 10:57 siC1_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 3830628 Dec 18 10:57 siC1_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 3941656 Dec 18 10:58 siC2_chr21_R2.fastq.gz
-rw-r--r--. 1 kkeith research 3078529 Dec 18 10:58 siC2_chr21_R2.fastq.gz
-rw-r--r--. 1 kkeith research 3078529 Dec 18 10:59 siC3_chr21_R1.fastq.gz
```

- Fastq files (usually) end in either fastq.gz or fq.gz (or they can be missing the .gz extension)
- File names will have some combination of the following information (depends on the sequencer):
 - Sample ID
 - Lane
 - Read number
 - Unique ID from the company or sequencer

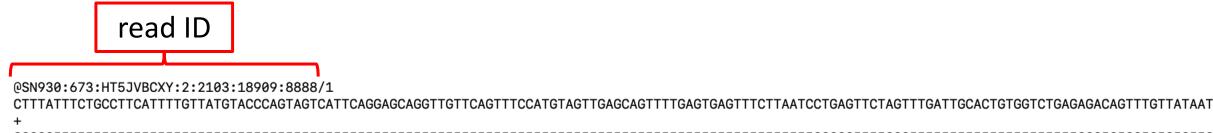
[[kkeith@cbix rnaseq_data]\$ 11

```
total 46700
```

```
-rw-r--r--. 1 kkeith research 4276449 Dec 18 10:55 dac1_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 4441834 Dec 18 10:55 dac1_chr21_R2.fastq.gz
-rw-r--r--. 1 kkeith research 4118184 Dec 18 10:55 dac2_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 4296786 Dec 18 10:55 dac2_chr21_R2.fastq.gz
-rw-r--r--. 1 kkeith research 4336091 Dec 18 10:56 dac3_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 4519748 Dec 18 10:56 dac3_chr21_R2.fastq.gz
-rw-r--r--. 1 kkeith research 3652875 Dec 18 10:57 siC1_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 3830628 Dec 18 10:57 siC1_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 3941656 Dec 18 10:58 siC2_chr21_R2.fastq.gz
-rw-r--r--. 1 kkeith research 3078529 Dec 18 10:58 siC2_chr21_R2.fastq.gz
-rw-r--r--. 1 kkeith research 3078529 Dec 18 10:59 siC3_chr21_R1.fastq.gz
-rw-r--r--. 1 kkeith research 3198840 Dec 18 10:59 siC3_chr21_R2.fastq.gz
```

```
[[kkeith@cbix Rawdata]$ ll [fm]*/*.fq.gz
```

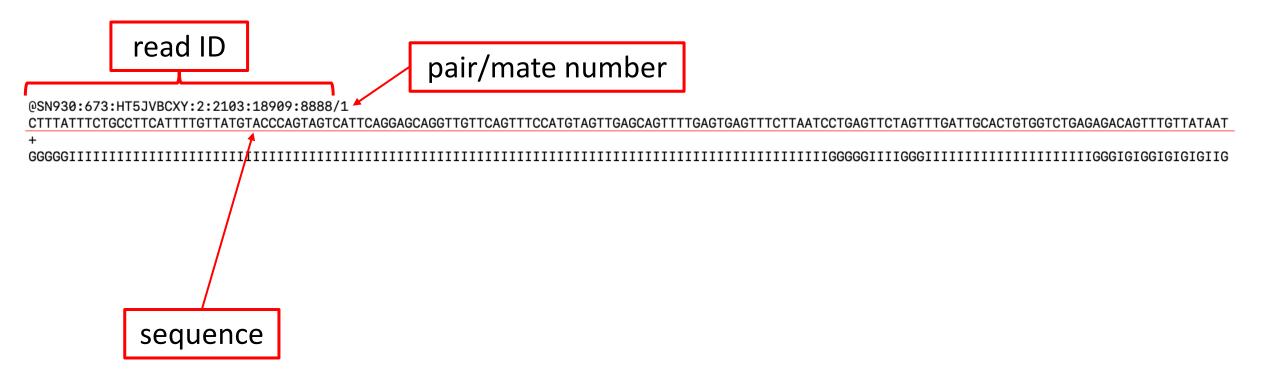
[[KKGICH@ODIX Kuwaucu]@	TT LINDAAAAAA	2			
-rw-rr 1 jjelinek	research 49946	4738 De	c 10	20:02	f45y4/f45y4_CKDL190143587-1a-6_H723FBBXX_L1_1.fq.gz
-rw-rr 1 jjelinek	research 53290	7223 De	c 10	20:03	f45y4/f45y4_CKDL190143587-1a-6_H723FBBXX_L1_2.fq.gz
-rw-rr 1 jjelinek	research 43784	.4462 De	c 10	20:01	f53y6/f53y6_CKDL190143587-1a-12_H723FBBXX_L1_1.fq.gz
-rw-rr 1 jjelinek	research 46882	.0826 De	c 10	20:02	f53y6/f53y6_CKDL190143587-1a-12_H723FBBXX_L1_2.fq.gz
-rw-rr 1 jjelinek	research 43534	4157 De	c 10	20:01	f61y8/f61y8_CKDL190143587-1a-19_H723FBBXX_L1_1.fq.gz
-rw-rr 1 jjelinek	research 45896	5705 De	c 10	20:01	f61y8/f61y8_CKDL190143587-1a-19_H723FBBXX_L1_2.fq.gz
-rw-rr 1 jjelinek	research 45953	9700 De	c 10	20:00	m38y1/m38y1_CKDL190143587-1a-2_H723FBBXX_L1_1.fq.gz
-rw-rr 1 jjelinek	research 49059	6742 De	c 10	20:01	m38y1/m38y1_CKDL190143587-1a-2_H723FBBXX_L1_2.fq.gz
-rw-rr 1 jjelinek	research 37453	1888 De	c 10	20:00	m45y3/m45y3_CKDL190143587-1a-5_H723FBBXX_L1_1.fq.gz
-rw-rr 1 jjelinek	research 39639	1106 De	c 10	20:00	m45y3/m45y3_CKDL190143587-1a-5_H723FBBXX_L1_2.fq.gz
-rw-rr 1 jjelinek	research 42372	8070 De	c 10	20:01	m53y5/m53y5_CKDL190143587-1a-7_H723FBBXX_L1_1.fq.gz
-rw-rr 1 jjelinek	research 44975	8575 De	c 10	20:02	m53y5/m53y5_CKDL190143587-1a-7_H723FBBXX_L1_2.fq.gz
-rw-rr 1 jjelinek	research 40612	8079 De	c 10	20:00	m61y7/m61y7_CKDL190143587-1a-16_H723FBBXX_L1_1.fq.gz
-rw-rr 1 jjelinek	research 43082	4343 De	c 10	20:01	m61y7/m61y7_CKDL190143587-1a-16_H723FBBXX_L1_2.fq.gz

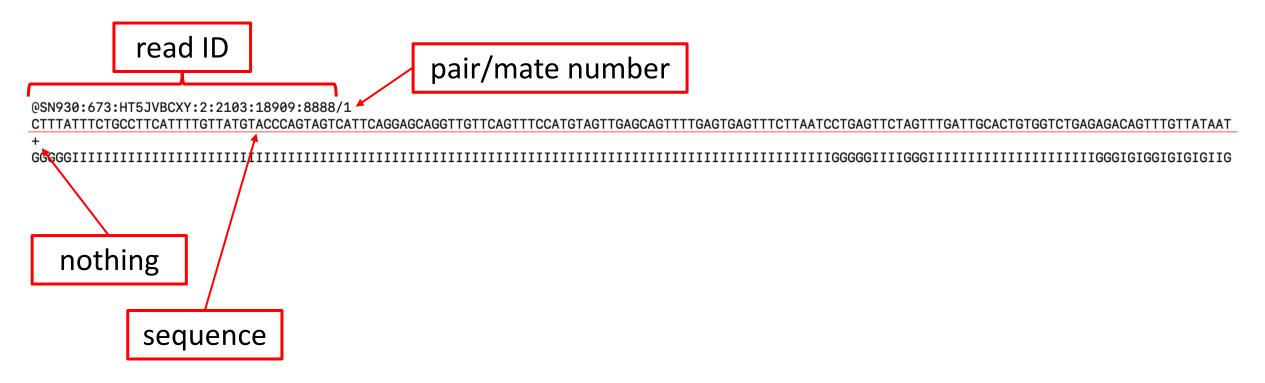




@SN930:673:HT5JVBCXY:2:2103:18909:8888/1

read ID









esn930:673:HT5JVBCXY: CTTTATTTCTGCCTTCATTT GGCGGGIIIIIIIIIIII nothing sequ

NOTE: YOUR SEQUENCING DATA WILL FREQUENCLY LOOK DIFFERENT

- File names vary from sequencer to sequencer
 - Read IDs also depend on the sequencer and will probably be different from the example here
 - Quality encoding can be different if you're using older or public data

with 0=unused, 1=unused, 2=Read Segment Quality Control Indicator (bold (Note: See discussion above).

GAGAGACAGTTTGTTATAAT

IIIGGGIGIGGIGIGIGIGIG

hijklmnopgrstuvwxyz{

126

L - Illumina 1.8+ Phred+33, raw reads typically (0, 41)

Quality Check

FastQC

- Before going forward, we want to check the quality of the data
 - How much did the sequencer fail?
 - Did we sequence mostly our sample DNA?
- FastQC is a program from the Babraham Institute in the UK that creates an html report on the quality of the sequencing data
 - Has 11 quality control checks that it does

Basic Statistics

Good Quality

Basic Statistics

Measure	Value
Filename	good_sequence_short.txt
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	250000
Sequences flagged as poor quality	0
Sequence length	40
%GC	45

Bad Quality

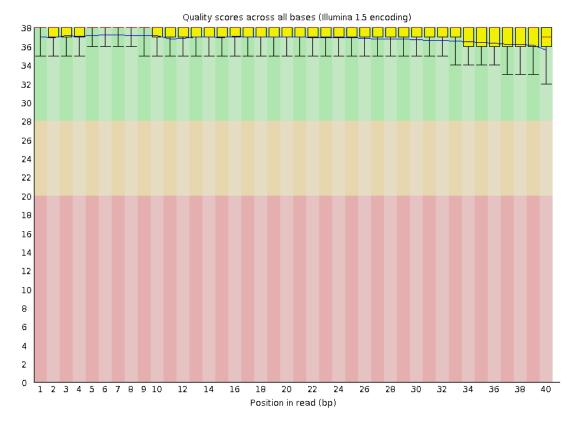
Basic Statistics

Measure	Value
Filename	bad_sequence.txt
File type	Conventional base calls
Encoding	Illumina 1.5
Total Sequences	395288
Sequences flagged as poor quality	0
Sequence length	40
%GC	47

Per base sequence quality

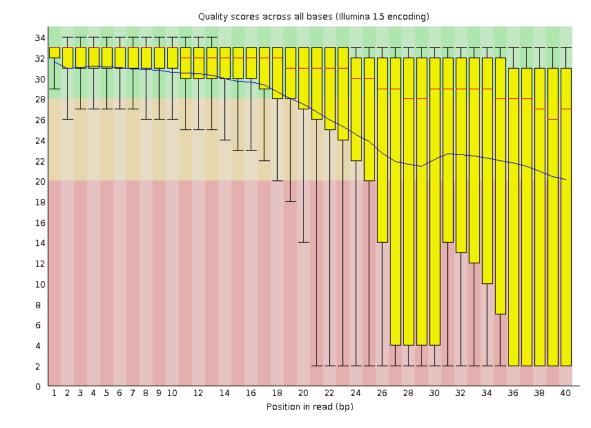
Good Quality

Per base sequence quality



Bad Quality

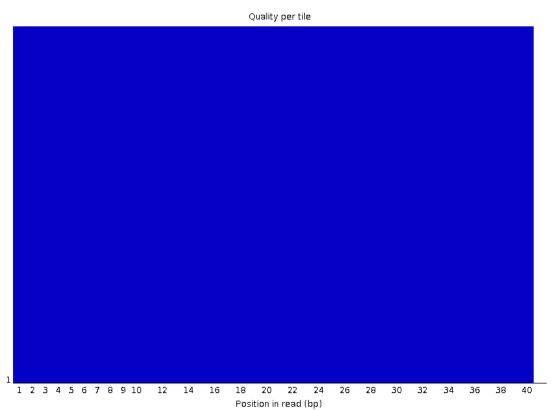
OPer base sequence quality



Per tile sequence quality

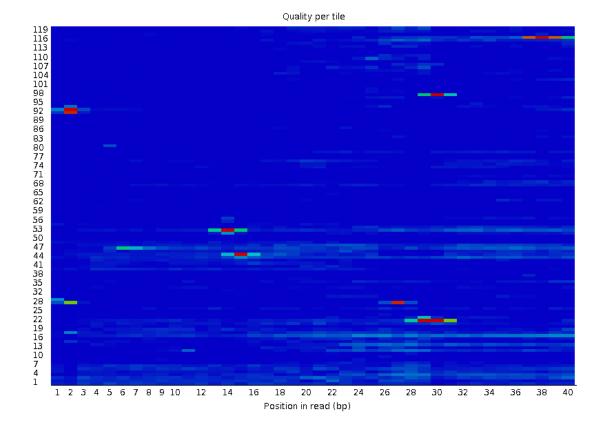
Good Quality





Bad Quality

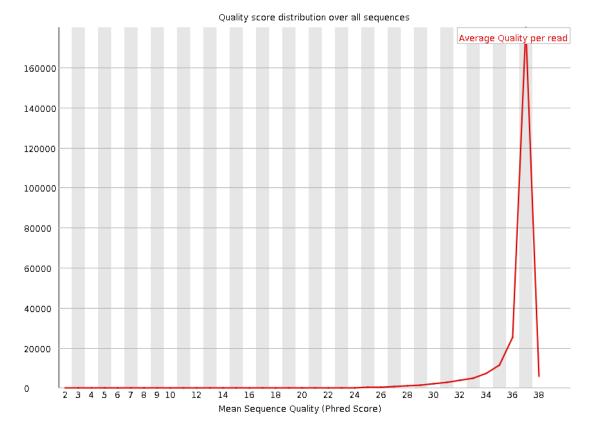




Per sequence quality scores

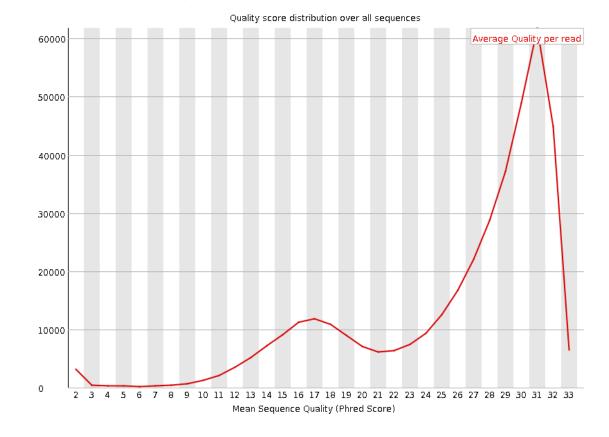
Good Quality

Per sequence quality scores



Bad Quality

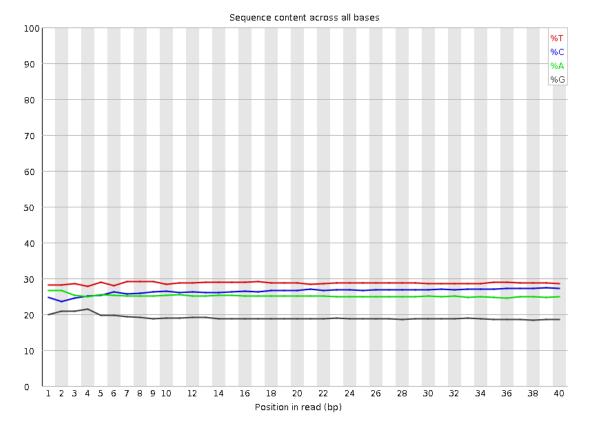
Per sequence quality scores



Per base sequence content

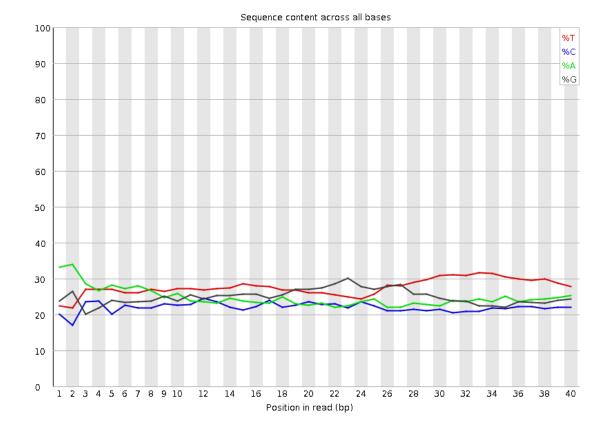
Good Quality

Per base sequence content



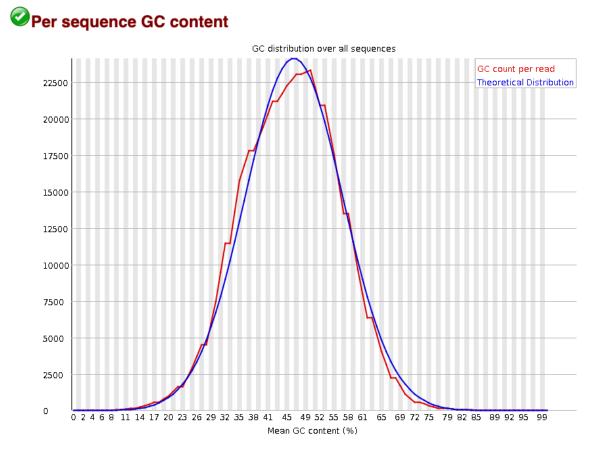
Bad Quality

UPer base sequence content



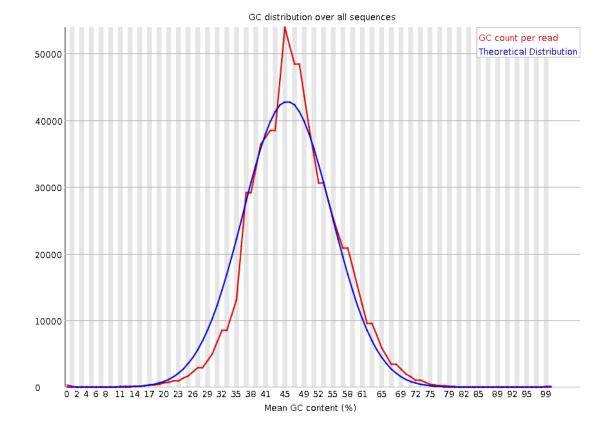
Per sequence GC content

Good Quality



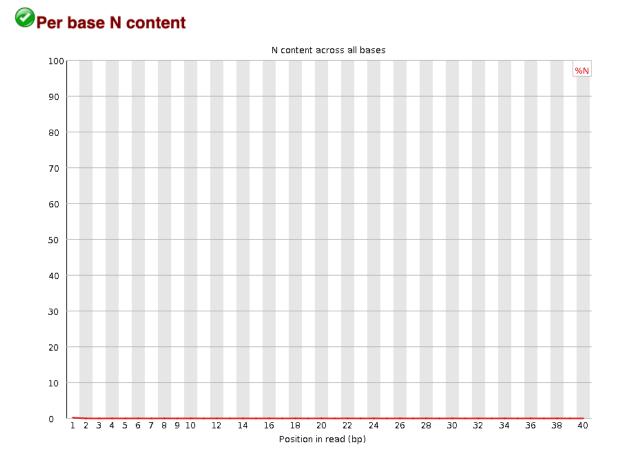
Bad Quality

Per sequence GC content



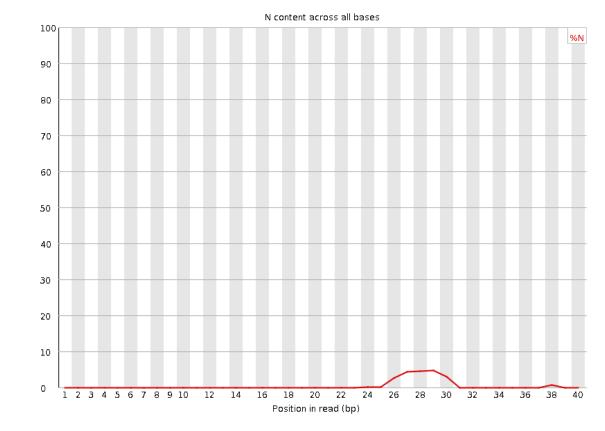
Per base sequence quality

Good Quality



Bad Quality

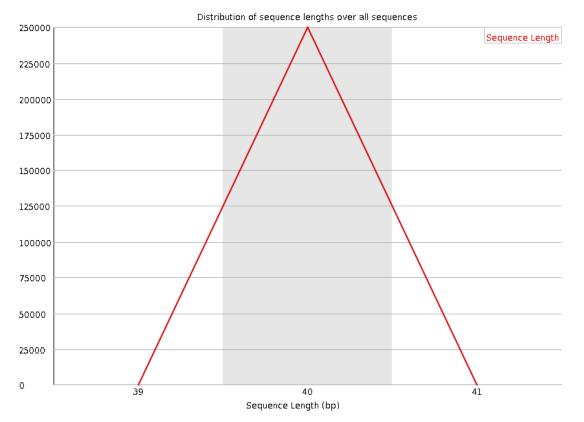
Per base N content



Per base sequence quality

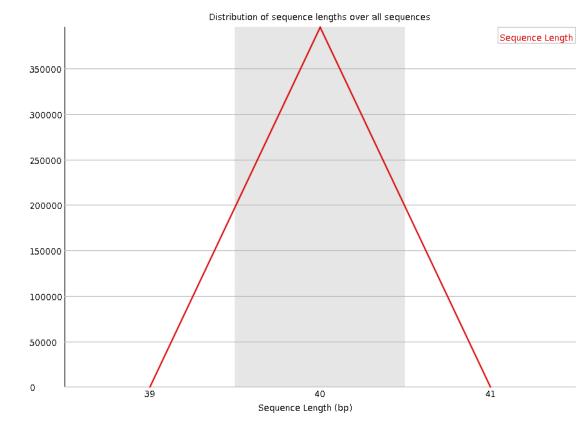
Good Quality

Sequence Length Distribution



Bad Quality

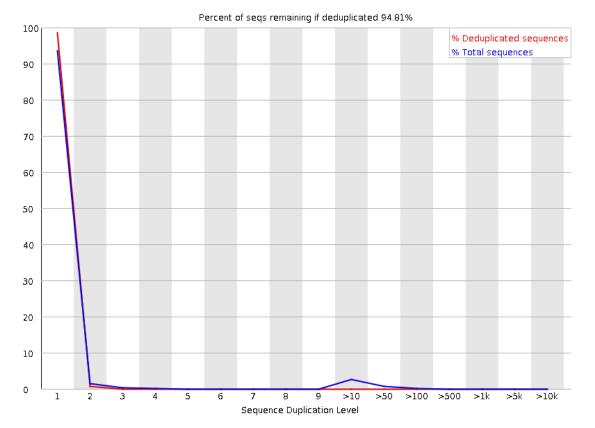
Sequence Length Distribution



Sequence Duplication Levels

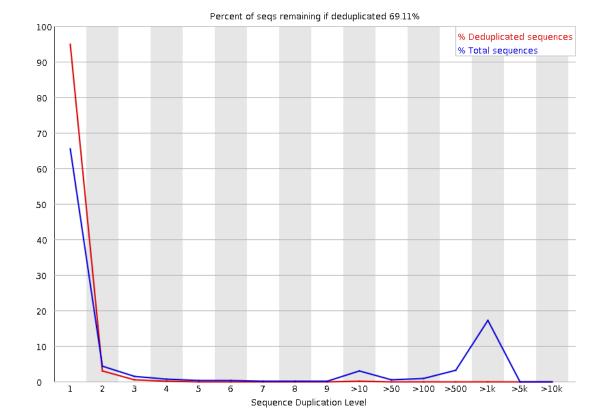
Good Quality

Sequence Duplication Levels



Bad Quality

Sequence Duplication Levels



Overrepresented sequences

Good Quality



No overrepresented sequences

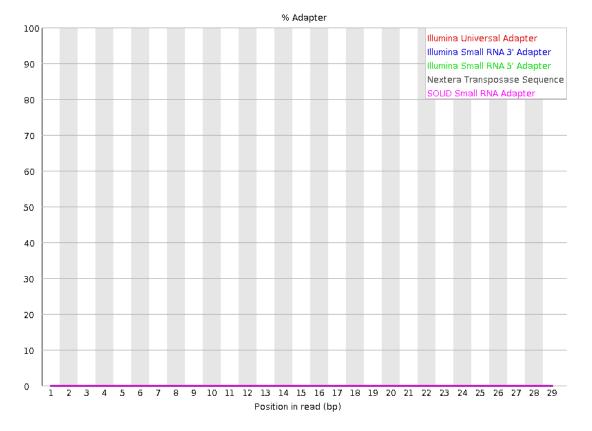
Bad Quality

NUMBER 2000 0.2234911143640 30 111 NUTROGONZACIONALMENTERNAMENTEN 2014 0.51319274249171 30 111 NUTROGONZACIONALMENTERNAMENTEN 2114 0.41319442914714 30 111 NUTROGONZACIONALMENTERNAMENTEN 1110 0.413194419149171 30 111 NUTROGONZACIONALMENTENTENTENTENTE 1114 0.447319171017171 30 111 NUTROGONZACIONALMENTENTENTENTENTENTE 1110 0.447319171017171 30 111 NUTROGONZACIONALMENTENTENTENTENTENTE 1210 0.4493149131117 30 111 NUTROGONZACIONALMENTENTENTENTENTENTENTENTE 1210 0.44931491311417 30 111 NUTROGONZACIONALMENTENTENTENTENTENTE 1210 0.41931491411417 30 111 NUTROGONZACIONALMENTENTENTENTENTENTENTENTENTENTENTENTENTE	Sequence	Count	Percentage	Possible Source
NTTOCONCENSIONALCTICUMUTTYRECONTECTION 1014 0.4355942879314 10 111 CMATAMATATIGUCTATCOMMUTTATION 1019 0.4355942879314 10 111 AMAATATIGUCTATCOMMUTTATION 144 0.4355942879313 10 111 AMAATATIGUCTATCOMMUTTATION 144 0.445931741801874 10 111 AMAATATIGUCTATCOMMUTTATION 144 0.443931741801874 10 111 AMATATIGUCTATCOMUTACINATIONALITY 113 0.443931741801874 10 111 AMATATIGUCTATCOMUTACINATIONALITY 113 0.443931741801874 10 111 AMATATIGUCTATCOMUTACINATIONALITY 113 0.44393174810174 10 111 AMATATIGUCTATCOMUTACINATIONALITY 113 0.43393174810174 10 111 AMATATIGUCTATCOMUTACINATIONALITY 113 0.43393174810174 10 111 AMATATIGUCTATCOMUTACINATIONALITY 114 0.43393174810174 10 111 CARACTECOMANTATICOUTCOMUTACINALITY 114	AGAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTTTC	2065		No Hit
CATABASINGTOCONSCICALCOCIALCITYN [1] 0.4439441584704004 60 11 AMANTGATUGOUTATCAACCTOCIALGUTTIATCOCT 144 0.4139441584704 80 11 AMANTGATUGOUTATCAACCTOCIALGUTTIATCOCT 144 0.41394115860714581 80 11 CURCUMACTUCACCTOCIALGUTTIATCOCTUCCUM 144 0.41394115860714581 80 11 CURCUMACTUCACCTOCIALGUTTIATCOCTUCCUM 144 0.41394115860714581 80 11 CURCUMACTUCACCTOCIALGUTTIATCOCTUC 170 0.439841586741631 80 11 ANANTATUGOUTACCAACCTOCIALGUTTIATCOCTUC 170 0.43984036893164 80 11 ANANTATUGOUTACCAACCTOCIALGUTTIATCOCTUC 170 0.4398403693164 80 11 ANANTATUGOUTACCAACCTOCIALGUTTIATCOCTUC 170 0.4398403693164 80 11 ANANTATUGOUTACCAACCTOCIALGUTTIATCOCTUC 170 0.4398403693164 80 11 ANANTATUTIACCTUCAACCTOCIALGUTACAACTUC 140 0.42919416131512 80 11 ANANTATUTIACCTUCAACCTOCIALGUTACAACTUC 140 0.4129194161321 80 11 CURCUMACTUCAACCTOCIALGUTACAACTUCAACTUC 140 0.4129194161321 80 11 CURCUMACTUCAACCTOCIALGUTACAACTUCAACTUC 140				
CARACTACTOCARCTECONACTTACCCTACACCONSTITUTACE1410.47501273017312No. NicCARACTACTOCARCTECONACTTACCCALACTTACC14300.4543147147343273No. NicACCTOCARCTECONACTCACACCONSTITUTACE13100.45431471473432733No. NicACCTOCARCTECONACTCACACCONSTITUTACE13100.45431471473432733No. NicALTACATTOCONTCCCALCTOCARACTTATACCONTCC17300.4583163743153147No. NicALTACATTOCONTCCANCTCONACTTATACCONTCC17400.4583163743153147No. NicALTACATTOCONTCCANCTCONACTTATACCONTCC17400.4393163231483147No. NicALTACATTOCONTCCANCTCONACTTATACCONTCC17400.43931632031483147No. NicCARCATCACCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCC17400.43931632031212No. NicCARCATCACCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANCTANCTOCANCTANCTA15400.43931459314276148No. NicCARCATCACATCACACAGAAAAATTATACCONTCCANCTANCANCTANCT15400.3983237631828No. NicNicCARCATCACATCACACAGAAAATTATACCONTCCANCTANCANCTANCTANCTANCTANCTANCTANCTA	ATTGGCGTATCCAACCTGCAGAGTTTTATCGCTTCCATGA	2014	0.5095019327680071	No Hit
CARACTACTOCARCTECONACTTACCCTACACCONSTITUTACE1410.47501273017312No. NicCARACTACTOCARCTECONACTTACCCALACTTACC14300.4543147147343273No. NicACCTOCARCTECONACTCACACCONSTITUTACE13100.45431471473432733No. NicACCTOCARCTECONACTCACACCONSTITUTACE13100.45431471473432733No. NicALTACATTOCONTCCCALCTOCARACTTATACCONTCC17300.4583163743153147No. NicALTACATTOCONTCCANCTCONACTTATACCONTCC17400.4583163743153147No. NicALTACATTOCONTCCANCTCONACTTATACCONTCC17400.4393163231483147No. NicALTACATTOCONTCCANCTCONACTTATACCONTCC17400.43931632031483147No. NicCARCATCACCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCC17400.43931632031212No. NicCARCATCACCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANACTTATACCONTCCANCTOCANCTANCTOCANCTANCTA15400.43931459314276148No. NicCARCATCACATCACACAGAAAAATTATACCONTCCANCTANCANCTANCT15400.3983237631828No. NicNicCARCATCACATCACACAGAAAATTATACCONTCCANCTANCANCTANCTANCTANCTANCTANCTANCTA				
NAMARTOGOURACCACTOCAMACTERATOR1440.44571012701723NoNiCENTODACTACCACTOCAMACTERATOR14410.445714719502875NoNiCATADACCTOCAMACTERATORCACTOCAMACTERATOR1310.445041719502875NoNiCATADATOGOURACCACACTECAMACTERATORCCTE1770.4085016974051814NoNiANARTATOGOURACCACCTECAMACTERATORCCTE1770.4085016974051814NoNiANARTATOGOURACCACCTECAMACTERATORCCTE1720.43704602653264NoNiCATADATOGOURACCACCTECAMACTERATORCCTE1720.43704602653264NoNiCATACTACACCTECAMACTERATORCCTECA1780.43250429522174NoNiCATACTACACCTECAMACTERATACCTE1740.432514927522174NoNiCATACTACACCTECAMACTERATACCTE1740.432514792722174NoNiCATACTACACCTECAMACTERATACCTE17400.4325172250221NoNiCATACTACACCTECAMACTERATORCCAMA17400.43951721722182NoNiCATACTACACCTECAMACTERATORCCAMA17400.43951721722182NoNiCATACTACACCTECAMACTERATORCCAMA17400.43951721722182NoNiCATACTACACCTECAMACTERATORCCAMACTERA17400.3395229502218NoNiCATACTACACCTECAMACTERATORCCAMACTERA17400.33952295022182NoNiCATACTACACCTECAMACTERATORCCAMACTERATORCAMACTERATORCCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORCAMACTERATORC				
TANTOCONTANCLAACTEGALANTTTACCONTON1410.44573475419995NoILIAACTOCAMOUTTACTOCATCAACTEGALANTTTACCONTON1330.445234573452735NoILIAATALTATOGOCTACCAACTEGALANTTTACCONTON1370.45631637413147NoILIAATALTATOGOCTACCAACTEGALANTTTACCONTON1370.45631637413147NoILIAATALTATOGOCTACCAACTEGALANTTTACCONTON1370.43534263453147NoILIAATALTATOGOCTACCAACTEGALANTTTACCONTON1310.43354263453147NoILICATACTEGALANTTACCONTECAATAGEGALANTTACT1440.4269191252214NoILICATACTEGALANTTACCONTECAATAGEGALANT1440.4269191252214NoILICATACTEGALANTTACCONTECAATAGEGALANT1440.42797814215212NoILICATACTEGALANTTACCONTECAATAGEGALANT1440.42797814215212NoILICATACTEGALANTTACCONTECAATAGEGALANTA1440.42797814215212NoILICATACTEGALANTTACCONTECAATAGEGALANTA1440.42797814215212NoILICATACTEGALANTTACCONTECAATAGEGALANTTAC1440.3798514512754NoILICATACTEGALANTTACCONTECAATAGEGALANTTACCONTECAATAGE1440.3798514512754NoILICATACTEGALANTTACCONTECAATAGEGALANTTAC1440.3798514512754NoILICATACTEGALANTTACCONTECAATAGEGALANTTACCONTECAATAGE1440.3798514512754NoILICATACTEGALANTTACCONTECAATAGEGALANTTAC1440.3798514512754NoILICATACTEGALANTTACCONTECAATAGEGALANTTAC1450.37981451				
AACTOCALCONCALCTOCALACTOCALACTICA AATAATAGOCATACCALCTOCALACTICALACTICAL AATAATAGOCATACCALCTOCALACTICALACTICAL BIE AATAATAGOCATACCALCTOCALACTICALACTICAL BIE AATAATAGOCATACCALCTOCALACTICALACTICAL BIE AATAATAGOCATACCALCTOCALACTICALACTICAL BIE AATAATAGOCATACCALCTOCALACTICALACTICAL BIE AATAATAGOCATACCALCTOCALACTICALACTICAL BIE AATAATAGOCATACCALCTOCALACTICALACTICAL BIE AATAATAGOCATACCALCTOCALACTICALACTICAL BIE AATACTAGOCATACCALCTOCALACTICALACTICAL BIE AATACTAGOCATACTICALACTICALACTICAL BIE AATACTAGOCATACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICALACTICAL BIE ACCALCTOCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACTICALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICAL BIE ACCALCALACTICALACTICALACTICAL BIE ACCALCALAC				
SHATAMATERTEGECTANCEGEMENTTATEGETE 131 0 4.35295142793555 No NE AMATAMATERTEGECTANCEGEMENTTATEGETE 179 0 4.352914079413147 No NE AMATAMATEGECTANCEGEMENTATEGETE 179 0 4.352434493534361 No NE AMATERTEGECTANCEGEMENTATEGETE 170 0 4.352434493534361 No NE AMATERTEGENTACCANCETEGANATEGETE 170 0 4.352434543534361 No NE CARACETEGANATETATEGETECCATAGEGEMENT 164 0 4.323953245745823 No NE CARACETEGANATETATEGETECCATAGEGEMENTATACE 164 0 4.21379816213221 No NE CARACETEGANATETATEGETECCATAGEGEMENTATACE 160 0 161 161 161 CARACETEGANATETATEGETECCATAGEGEMENTATACE 161 0 161				
AAATGADCONSCOLANCTTATACCONTEN17790.4550140794131147No. III.ANTACTATOGCONTACCONCONCANTTATACCONTEN17500.45521407450131147No. III.AAATATATOGCONTACCONCONCONTTATATOCONTEN17500.45521407450145147No. III.AAATATATOGCONTACCONCONCONTTATATOCONTEN17500.4352140940145148No. III.AAATATATATOGCONTECCANAGOGANAGT17600.43531407401451218No. III.ATCOANCETCONAGOTTATATOCONTECCANAGOGANAGT1660.42519181251218No. III.CACACTOTAGADATATATACONTECANAGOGANAGTATALACI1660.42519181251218No. III.CACACTOTAGADATATATACONTECANAGOGANAGTATALACI1660.42519181251218No. III.CACACTOCAAGATATATACONTECANAGOGANAGTATACONTE1600.435237157251218No. III.CACACTOCAAGATATATACONTECANAGOTACONTACACTO1500.395237157121828No. III.CACACTOCAAGATATATACONTECANAGOTCONTACATOCONTECANAGO15100.395237157121828No. III.CACACTOCAAGATATATATACONTECANAGOTCONTACACTOCONTECANAGOT15100.395237157121828No. III.CACACTOCAAGATATATATATATATATATATATATATATATATATAT				
AMARTANTOGOTARCOAACCTGOADAGTTTARCOCTTCO AATGATGOTGOTARCOAACCTGOADAGTTTARCOCTTCO TFG 0 4452498593461 No Hit AATGATGOTGOTARCOAACCTGOADAGTTTARCOCTTCO TFG 0 4452498593461 No Hit CARACTGATGOGTARCOAACCTGOADAGTTARCOCTTC 10 6.3339802498128 No Hit CARACTGATGOGTARCOAACGTARCOACTAA H 4 0.434918439128 No Hit CARACTGCOAGTTCOATGATGOCTACCARLOCOAD 11 6.3339802498128 No Hit CARACTGCOAGTTCOATGATGOCTACCARLOCOAD 14 0.43291912412912 No Hit CARACTGCOAGTTCOATGATGOCTACCARLOCOAD 14 0.43291912412912 No Hit CARACTGCOAGTTCOATGATGOCTACCARLOCOAD 14 0.43291912412912 No Hit CARACTGCOAGTTCOATGATGOCTACCARLOCOAD 14 0.43291912412912 No Hit CARACTGCOAGTTCOATGATGOCTACACTGCOA 14 0.43291912412912 No Hit CARACTGCOAGTTCOATGATGOCTACACTGCOA 14 0.43291912412912 No Hit CARACTGCOAGTTCAACTGCOATGATGOCTAC 14 0.43291912412912 No Hit CARACTGCOAGTTCAACTGCOAGTTCAA 14 0.439993229391645142975 No Hit CARACTGCOAGTTCAACTGCOAGTTTTACC 14 0.39993129731489 No Hit CARACTGCOAGTTCAACTGCOAGTTCAACTGCOA 14 0.3999312973149174 No Hit CARACTGCOAGTTCAACTGCOAGTTTTACC 14 0.3999312973149174 No Hit CARACTGCOAGTTCAACTGCOAGTTTTACC 14 0.3999312973149174 No Hit CARACTGCOAGTTCAACTGCOAGTTTTACC 14 0.3999312973149174 No Hit CARACTGCOAGTTCAACTGCOAGTCAATTTTAC 14 0.39993129731491792 No Hit CARACTGCOAGTTCAACTGCOAGTCAATTTTAC 14 0.39993129731414218 No Hit CARACTGCOAGTTCAACTGCOAGTCAATTTTAC 14 0.39993129731491 No Hit CARACTGCOAGTTCAACTGCOAGTCAATTTTAC 14 0.399931297314418 No Hit CARACTGCOAGTTCAACTGCOAGTCAATTTTAC 14 0.3999409731349 No Hit CARACTGCOAGTTCAACTGCCAATCATTTTAC 14 0.3993049731490 No Hit CARACTGCOAGTTCAACTGCCAATCGCCAATCATTTTAC 14 0.3993049731397 No Hit CARCTGCOAGTTCAACTGCCAATCATTTTAC 14 0.3993049731297 No Hit CARCTGCOAGTTCAAACTGCCAATCATTTTAC 14 0.3993049731297 No Hit CARCTGCOAGTTCAAACTGCCAATCATTTTAC 14 0.3993049731297 No Hit CARCTGCAAGTTCAATCGCCAATCATTTTAC 14 0.3993049				
AND ATCONCIDENCE CONCADULT TATECONTY 179 0.43742498934304 No NIL AMARYINITOGONAL TATACCONTY 179 0.43742462693434 No NIL ATCANCECTOCAMACTITATACCONTICATINGONAL 178 0.43354926984164 No NIL ATCANCTITATACCONTICATINGONALMACTITATACUT 0.44614897083474 No NIL CACADITITATACCONTICATINGONALMATTALACUT 0.4461489704724 No NIL CACACITITATACCONTICATINGONALMATTALACUT 0.4461497047275 No NIL CACACITITATACCONTICATINGONALMATTACCONTICATINGON 0.41092371217147726 No NIL CACACITICATINGONTICATIN				
AAAMATCHICOGONGAUTTATACOUT1719000<				
OCHAZALCEDCAGADETTITATCOCTTCCATAGOCIADA 1718 0.4325990284079253 No NIL ATSCAGACTCACAGADETTITATCOCTTCCATAGOCIADACTTA 164 0.4269109264079253 No NIL CACADITTITATCOCTTCCATAGOCIADACTTA 164 0.4269109264079253 No NIL TCCAADITTITATCOCTTCCATAGOCIADACTTA 164 0.42197016213212 No NIL CACATCICAADITTITATCOCTTCCATAGOCIADACTTA 164 0.43197016213212 No NIL CACATCICAADITTITATCOCTTCCATAGOCIADACTA 163 0.43197016213212 No NIL CACATCICAADITTITATCOCTTCCATAGOCIADACTA 164 0.4089179711420724 No NIL CACATCICAADITTITATCOCTTCCATAGOCIA 164 0.4989189714992184 No NIL CACATCICAADITTITATCOCTTCCATAGOCIANATITITATCOCTCCATAGOCIA 163 0.375175191298184 No NIL CACATCICAADITTITATCOCTCCATAGOCIANATITITATCOCTCCATAGOCIA 163 0.375175191298184 No NIL CACATCICAADITTITATCOCTCCATAGOCIANATITITA 164 0.375179139127618 No NIL CACATCICAADITTITATCOCTCCATAGOCIANATITITA 163 0.37521791277618 </td <td></td> <td></td> <td></td> <td></td>				
ACCADACTICACCONCRUCTCACANCOCCUANCY 199 0.1299020407923 No NI CAMACTICACCONCRUCTCACANCOCCUANCYTATACACTIC 0.0401197014235124 No NI CAMACTICACCONCRUCTCACANCOCCUANCYTATACACTICACUANCACUAN 0.0412107321205321 No NI CAMACTICACCONCRUCTCACANCOCCUANCACUAN 160 0.412307321205321 No NI ACCANCONCRUCANCONCUNCATICACUANCACUAN 160 0.4881987213205321 No NI ACCANCONCUNCANCONCUNCATICACUANC 160 0.4881987213205321 No NI ACCITACONCUNCANCONCUNCATICACUANC 120 0.3999323958428 No NI NI ACCITACONCUNCANCONCUNCATICACUANC 120 0.39993239582428 No NI NI ACCITACONACUCANCANCONCUNCATICACUANC 120 0.399931995824828 NI NI NI ACCITACONACUCANACUCANCONCUNCAUNCAUNCAUNCAUNCAUNCUNCAUNCUNCAUNCUNCAUNCUNCAUNCUNCAUNCUNCAUNCUNCUNUUUUUUUU				
CALANTETTATCONTOCATIONCORDINATIONCONT 144 0.4210701225124 No No CONCATTETATOCONTOCATIONCORDINATIONCO 160 0.42107012512124 No No CARCITECAMANTETATOCONTOCATIONCORDINATION 160 0.42107012512124 No No CARCITECAMANTETATOCONTOCATIONCORDINATION 160 0.42107012712714 No No CARCITECAMANTETATOCONTOCATIONCONTO 161 0.498279111471471714 No No CARCITECAMANTETATOCONTOCATIONCONTO 164 0.3989356312419714 No No CONTANTATOTOCAMANTETATOCONTOCATIONCONTO 164 0.398935787542892 No No No CONTANTANTETOCONTOCATIONCONTOCATIONCONTO 164 0.39893787542892 No No No CONTANTANTETOCONTOCATIONCONTOCATIONCONTO 164 0.39893787542892 No No No CONTANTANTOTOCONTOCATIONCONTOCATI				
TOCAMACTTYANCOCTTOCATAGOCGUMANTTA 164 0.41179161235124 No Ni CAMACTGOCAMACTTYANCOCTTOCATAGOCGUMANTTA 164 0.4112915125128 No Ni CAMACTGOCAMACTTYANCOCTCOCATAGOCGUMANTTA 164 0.4112915125128 No Ni CATAGOMACTUCAUGUCGUMATAGOCGUMANT 164 0.4112915127143714 No Ni CATAGOMACTUCAUGUCGUMATAGOCGUMAT 164 0.4112915127143714 No Ni CATAGOMACTUCAUGUCGUMATAGOCGUMAT 124 0.3199532950282 No Ni Ni CATAGOMACTUCAUGUTGUAUGUTAUGUCGUMATTATA 124 0.3199532950282 No Ni Ni Ni CATAGOMACTUCAUATTATOCOTTCOCATAGOC 124 0.3199532950282 No Ni Ni Ni CATAGOMACTUCAUATTATOCOTTCOCATAGOCGUMATTA 124 0.319931790131787 Ni Ni Ni CATAGOMACTUCAUATTATOCOTTCOCATAGOCGUMATTA 124 0.3199241871 Ni Ni Ni CATAGOMACTUCAUATTATOCOTTCOCATAGOCGUMATTA 124 0.31992418714276 Ni Ni Ni CATAGOMACTUCAUATTAT				
CALCTOCKARACTTICSCONAUCCONAUCCON 144 0 0.41125757220521 0 111 CACKARGENGARMANCTTICSCONAUCCONA 143 0.411257572205221 0 111 CACKARGENGARMANCTTICSCONTIGGARACCONA 143 0.419237711447712 0 111 CACINGARGENTAMANCTTICSCONTIGGARACCONA 143 0.3492379171447712 0 111 CARGENGARGENTAMANCTTICSCONTIGGARACCONA 143 0.3492379173429714 0 111 CORONTICALACTOCARAGENTATACCONTIGA 143 0.3492379173429714 0 111 CARGENGARGENTAMACTOCICARAGENTATACCONTIGA 143 0.3492379173429714 0 111 CARGENGARGENTAMACTOCICARAGENTATACCONTIGA 143 0.349237917317574 0 111 CARGENGARGENTAMACTOCICARAGENTATACCONTIGA 143 0.35921791017912754 0 111 CARGENGARGENTAMACTOCICARAGENTAMACTOCICARAGENTATA 143 0.3492491245424 0 111 CARGENGARGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICARAGENTAMACTOCICAR				
TATCCANCIDANTTATCONCTOCANCOUND 1410 0.40923772140726 No Nici CACATOGANCIDATAMATCTOCANCITOCANCOCCAN 1420 0.40923772140726 No Nici CACATOGANCIDATAMATCTOCANCITOCANCOCCAN 1430 0.40923772140726 No Nici CALANCITATATCOCATCANCITOCANCITOCANCIDA 1430 0.3992539524922 No Nici CALANCITATATCOCATCANCINCANCIDA 1430 0.399253952395228 No Nici CANCITATATCOCATCOCANCINCANCICANCINCANCIDA 1430 0.39925395395228 No Nici CANCITATATCOCATCOCANCICANCINCANCICANCINC				
Gramodakocaraakacretockoortoaktaccoakt Acteressortandekokakakacretokkoortakakart Golakortakacretokakoitakakart Golakortakacretokakoitakakart Golakortakacretokakoitakakart Golakortakakart Gol				
AACTYCOGCAGACGUCAGAGAGATAGACCTOCCAGAGAGTAGACAGA CAGAGAGTATAGCCTCCAGAGAGTAGACAGAGAGAGAGAG				
BORDADETTARCOCCTACUCACUCACUTACUTACUCA 1910 0.39939976322975428 No ILI GORDATACCACUCACUTACUTACUCACUTACUTACUCA 142 0.39939372612828 No ILI GORDATACCACUCACUTACUTACUCACUTACUTACUCA 141 0.39849372612828 No ILI ATAMAATTAGUCACUTACUTACUTACUCATURACUCA 141 0.397372512828428 No ILI ATAMAATTAGUCACUTACUTACUCACUTACUTAL 142 0.3973773137898274 No ILI ATAMAATTAGUCACUTACUTACUTACUCAUTAUT 142 0.397377310718728 No ILI ATAMAATTAGUCAUTAUAUTAUTAUTAUTAU 0.39737370187281848 No ILI COTAVAAUAUTAUAUTAUAUTAUAUTAUAUTAUAUTAUTUAUUTAU 0.39737370187281848 No ILI COTAVAAUAUTAUAUTUAUUTAUAUTUTUUUUUUUUUUUUU				
PROFECTACCONCLINETTATICCONTECCATIONE 154 0.39823785742982 No. NIC GROTATICCAACCTONAMITTATICCONTECCATIONE 154 0.3982837287342984 No. NIC ACTOCAAMITTATICCONTECCAACCTONAMIC 154 0.398283728734284 No. NIC ACTOCAAMITTATICCONTECCAAMITTATICCONTECCAAMIT 124 0.39721373742482384 No. NIC ACTOCAMITTATICCONTECCAAMITTATICCONTECCAAMIT 124 0.397213731382442 No. NIC CANTAMATCTOCAMITTATICCONTECCAAMITTATIC 124 0.397213731372238 No. NIC CANTAMATCTOCAMITTATICCONTECCAAMITTATIC 134 0.39722449731427 No. NIC CATTOTOCAMITTATICCONTECCAAMITTATIC 133 0.31722449731427 No. NIC CATTOTOCAMITTATICCONTECCAAMITTATIC 133 0.31722449731427 No. NIC CANTAMATCTOCAMITTATICCONTECCAAMITTATIC 134 0.3192447234586 No. NIC CANTAMATCTOCAMITTATICCONTECCAAMITTATIC 134 0.3192447234586 No. NIC CANTAMATCTOCAMITTATICCONTECCAAMITTATICCONTECCAAMITTATIC 134 0.3192447234586 No. NIC CANTAMATCTOCAMITTATICCONTECCAAMITTATICCONTECCAAMITTATICONTECAAMITTATICONTECCAAMITTATICONTECCAAMITTATICONTECAAMITTATICONTECAAMITTATICONTECAAMITTAT				
Georgenerocalact	GCAGAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTT	1580	0.39970856691829754	No Hit
ATAMAANCUNTUSCONTACCAACCTOCAACTUTTATCO ACCOUNTERATTATACCTOCAACTUTACCOUNTACCCCAATCA 1479 0.745257574342384 No HL CATCOUNTERATACTUTCAACTUTACCAACTUTATAC 1470 0.75527574342384 No HL CATCOUNTERAACTUTCAACTUTACCAACTUTATAC 1470 0.75527574342384 No HL CATCOUNTERAACTUTCAACTUTCAACTUTATAC 1480 0.357279131234274 No HL CATCOUNTERAACTUTCAACTUTCAACTUTATAC 1490 0.34527914324475 No HL CATCOUNTERAACTUTCAACTUTCAACTUTATAC 1310 0.3452819374475 No HL CATCOUNTERAACTUTCAACTUTCAACTUTATA 1330 0.325285939486164 No HL CATCOUNTERAACTUTCAACTUTCAACTUTATA 1330 0.325285939486164 No HL CATCOUNTERAACTUTCAACTUTCAACTUTATA 1330 0.325285939486164 No HL CATCOUNTERAACTUTCAACTUTATACTUTATA 1330 0.335285939486164 No HL CATCOUNTERAACTUTCAACTUTATA 1330 0.338548458632356 No HL CATCUTCAACTUTCAACTUTATATACAACTUTATA 1330 0.38564839532958 No HL CATCUTCAACTUTCAACTUTATATACAAACTUTATA 1340 0.3851649359328 No HL CATCUTCAACTUTCAACTUTATATACAAACTUTATA 1340 0.3851649359328 No HL CATCUTCAACTUTAACTUTATATATACAAACTU 1340 0.345164948542858 No HL CATCUTCAACTUTAACTUTAACTUTATATATATAA 1340 0.3451649459328 No HL CATCUTCAACTUTAACTUTAACTUTATATATATAA 1340 0.335721640451532 No HL CATCUTAACTUTAACTUTAACTUTAACTUTAA 1340 0.335721640451532 NO HL CATCUTAACTUTAACTUTAACTUTAACTUTAACTUTAA 1340 0.335721640451532 NO HL CATCUTAACTUTAACTUTAACTUTAACTUTAA 1340 0.335721640451320 NO HL CATCUTAACTUTA	TGGCGTATCCAACCTGCAGAGTTTTATCGCTTCCATGACG	1569	0.3969257857562082	No Hit
ACTOROMONITAMACTOCIONINGOLAMACTIVA (19 0.)4513776428034 No Hit ATGUMAGCHTAMACTOCIONINGOLAMACTIVITATO (20 0.)5931753013197 No Hit CATAMACTOCIONINGOLAMACTIVITATO (20 0.)5930175010329 No Hit ATTOCIONINGOLAMACTIVITATOLOGUTIALIA (20 0.)20070010020 ATACTIVICIONINGOLAMACTIVICIONINGUTIALIA (20 0.)20070010020 ACTOROMONINATAMACTOCIONINGUTIALIA (20 0.)20070010020 (20 0.)2007001 (20 0.)2007	GGCGTATCCAACCTGCAGAGTTTTATCGCTTCCATGACGC	1542	0.39009532290380683	No Hit
ACCURACESANANCTETECADOTTCARTACCECANCAT ACCURACESANANCTETECADOTTCARTACCECANCAT ACTA ANALCTETECADOTTCARTACCECANCE ACTA ANALCTETECADOTTCARTACCECANCE ACTA ANALCTETECADOTTCARTACCECANCE ACTA ANALCTETECADOTTCARTCCCANCE ACTA ANALCTETECADOTTCARTCCCANCERT ACTA ANALCTETECADOTTCARTCCCANCERT ACTA ANALCTETECADOTTCARTCCCANCERT ACTA ANALCTETECADOTTCARTCCCANCERT ACTA ANALCTETECADOTTCARTCCCANCERT ACTA ANALCTETECADOTTCARTCCCANCERT ACTA ANALCTETECADOTTCARTCCCANCERT ACTA ANALCTETECADOCTCCANCERT ACTA ANALCTETECADOTTCARTCCCANCERT ACTA ANALCTETECADOCTCCANCERT ACTA ANALCTETECADOCTCANCERTATTTARTCCAN ACTA ANALCTETECADOCTCANCERTATTARTCCANCE ACTA ANALCTETECADOCTCANCECANCERT ACTA ANALCTETECADOCTCANCECANCECANCERT ACTA ANALCTETECADOCTCANCECANCECANCERT ACTA ANALCTETECADOCTCANCECAN	ATAAAAATGATTGGCGTATCCAACCTGCAGAGTTTTATCG	1481	0.37466353645949285	No Hit
SATAMASCTTCOCADOTTOGATAGOCCARCATTTITAT 142 0.353217601371022 No. ILI. COTTATCOMOCTANANCTCOCADOTOMASCOCCA. 143 0.346276785216524 No. ILI. CATCTTGOCTATGOAAGCARAAACTCOCADOTOMASCOCA. 143 0.34667785216524 No. ILI. CATCTTGOCTATGOAAGCARAAACTCOCADOTOMASCOCA. 143 0.348880377604174 No. ILI. CATACTTCOCOTCATGOAAGCARAACTCOCAMCCATT 143 0.31922488037804174 No. ILI. COTATACAACTCOCAACTGOCAMCCATT 124 0.339584980458644 No. ILI. COTATACAACTCOCAACTGOCAACCATTTATT 124 0.3395849804580454 No. ILI. COTATACAACTCOCAACTGOCAACCATTTATTOCOCTCAACGACCA 124 0.3395849804592058 No. ILI. COTATACAACTCOCAACTGOCAACTATTTATTATCOCAACT 128 0.349524495732054 No. ILI. COTATACAACTGOCAACTATTTATACCAACTATT 128 0.349524495732054 No. ILI. CATACTTCOCAACTGOCAACTATTTATACCAACTATT 128 0.3495247952054 No. ILI. CATACTCOCAACTGOCAACTATTTATACCAACTATTTATACCAACTA 128 0.3495247952154 No. ILI. CATACTCOCAACTGOCAACTATTTATACCAACTATTTATACCAACTA 128 0.3495247454129854 No. ILI.	ACCTGCAGAGTTTTATCGCTTCCATGACGCAGAAGTTAAC	1479	0.37415757624820384	No Hit
OCTATOGINACCITANANCTCTOORDOCTOGENECOCCA 141 0.35727999170922 No Nit ARTTCTOORDOCTANGUAGCUATANANCTCTCONDOCTOG 134 0.35727999170922 No Nit ARTTCTOORDOCTANGUAGCUATANANCTCTCONDOCTOG 134 0.3572246924168 No Nit CATOGINACCUATANANCTCTCONDOCTOGATAGUCANTT 133 0.3722246924168 No Nit CATOGINACCUATANATCTCTCONDOCTOGATAGUCANTT 137 0.3258551959021774 No Nit TANANTANTOGUCATCCANCTCANCACUANTTTAT 134 0.33584093233486 No Nit CONTANCANCTOROLONTTATATATAGUAGUANTTATA 0.3358409323486 No Nit CONTANGUARCUATANANCTCTCOLONTTATATATAGUANCU 131 0.34564094933266 No Nit ANACCTCTOLONTOGATAGUCANCANTTTATATOGUANCU 131 0.34564094933266 No Nit ANACCTCTOLONTOGATAGUCANCANTTTATATOGUANU 131 0.34564094933266 No Nit ANACCTCTOLONTOGATAGUCANCANTTTATATOGUANU 131 0.34564094933267 No Nit ANACCTCOLONTOGATAGUCANCANTTTATATOGUANU 131 0.3456409433268 No Nit	ATGGAAGCGATAAAACTCTGCAGGTTGGATACGCCAATCA	1452	0.3673271133958026	No Hit
ACTCTCOCCCCCACCOLAGACCTALAACCTCCCCCCCCCCCCCCCCCCCCCCCCCC	GATAAAACTCTGCAGGTTGGATACGCCAATCATTTTTATC	1420	0.35923175001517876	No Hit
TAMETTERCONTENTIONALANCETERCHOTTO 1313 0.347124240824108 No No No CARGIMANDARTENCONDENTIMATENCE 1333 0.33722440824108 No No No CARGIMANDARTENCONDENTIMATENTIMAT 1333 0.33722440824108 No No No CARGIMANDARTENCONDENTIMATENCENCANCENTITAT 1333 0.33722440824108 No No No CARGIMANDARTENCONDENTIMATENCENCANCENCIALIZA 0.337263835969080144 No No No No CONTRECONDENTITATENCENCANCENCIALIZA 0.337224407239548 No	CGTCATGGAAGCGATAAAACTCTGCAGGTTGGATACGCCA	1412	0.3572079091700229	No Hit
CANDIALACTECTICADOTTORIARAGECTANT 133 0.31222440934109 No Ni CANDAMACTECTICADOTTORIARAGECTANTTATA 134 0.323855945040174 No Ni CANDAMACTECTICADOCOTTARATATATATA 134 0.323855945040174 No Ni CANDAMACTECTICADACTECANAGETATATATA 134 0.3336559450801644 No Ni CANDAMACTECTICADACTECANAGETATATATA 134 0.3336559450801644 No Ni CONTANCIGATICADACTECTICADACETATI 134 0.333652934588 No Ni CANDECTICADACTECTICADACETATITITATA 0.3336545932936 No Ni Ni CANDECTICADACETACONTANATTITATATECANA 133 0.36624459532936 No Ni ACTECTICADAGETAGEANCOCANTATTITATATECANA 133 0.36624459532936 No Ni ANACTECTICADAGETAGEANCOCANTATTITATECANA 134 0.3616609737103074 No Ni ANACTECTICADAGETAGEANCOCANTATTITATECANA 147 0.333717189369375 No Ni ANACTECTICADAGETAGEANCOCANTETTITATECANA 147 0.333717189369375 No Ni CATEGE	ACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTTGGA	1368	0.34607678452166524	No Hit
CGATALAMECTCTGLOGGTTGGATAGCOCANCGATTITTAT 134 0.328840037604174 106 111 TAMAMETTGGGGTAGCOCANCGACHARTITTTATC 127 0.320853940940144 106 111 CGATACCAACCOMCANCATTACCOMCANCACCA 128 0.3136409313408 10 112 CGATACCAACCOMCANCATTATCACCAMCACCA 123 0.31344702396388 112 CGATACCAACCOMCANCATTATTATCACAMCACCA 123 0.31344702396588 112 CAGCATAMAACTCTCOCAGTAGTACCCAACCATTT 113 0.3856420352095 10 112 AAGCCTGAAGATAMACCOCAACCATTTATTATCACAMAC 133 0.381542039262995 10 112 AAGCCTGAAGATAMACCOCAACCATTTATTATCACAMAC 133 0.381542039262995 10 112 AAGCCTGAAGATAMACCOCAACCATTTATTATCACAMAC 133 0.38154203926395 112 112 AAGCCTGAAGATAMACCTCGAAGATTTATACCAMAC 133 0.3815480373207 10 112 ATAMACTGCGAAGATATAACCACAACATTTATATCAMAC 133 0.3121773774 112 1231777457774 112 1231777457774 112 1231777457774 112 12317774774 112 1231777457774 <t< td=""><td>TAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTTG</td><td>1363</td><td>0.34481188399344276</td><td>No Hit</td></t<>	TAACTTCTGCGTCATGGAAGCGATAAAACTCTGCAGGTTG	1363	0.34481188399344276	No Hit
TAMAATCATGGCGTATCCAACCTCCAALGUTTTATCCC 127 0.323555930002144 No NL GCGTATCCGACCTGCAACGUTCCAALGUCGT 128 0.31244593333464 No NL GCGTATCCGACCTGCAACGUTCCAALGUCGTAAGTCCAA 128 0.3124459333464 No NL GGGACGUACCTGCAACGUTCGAACGUCCAATCGTT 128 0.39202448717897 No NL GGGACGUATAAACTCTCCCAAGUCCAATCGTT 128 0.39202448717897 No NL ACCTCCGACGUTGGAACGUCGATCGTTTTATCGCGAGUCGTTGT 128 0.382642459532936 No NL ACCTCCGACGUTGGAACGUCGAATCGATTTTATCGCGAGU 133 0.386242459532936 No NL AAACTCTCGACGUTGGAACGUCCAATCGATTTTATCGCGAGU 130 0.3862493732307 No NL AAACTCTCGACGUTGGAACGUCCAATCGATTTTATCGCGAGU 111 0.381566937310374 No NL ATAMACTCGCAGUTGGAACGUCCAATCGATTTTATCGCGAGU 111 0.38156493523174 No NL CTCGCCGUTGGAACGUCAATCGCCAATCGATTTTATCGCCGAGU 120 0.339711604533142 No NL CTCGUCGUTGGAACGUCAATCGUTGGAATCGCC 120 0.339711604533142 No NL	CATGGAAGCGATAAAACTCTGCAGGTTGGATACGCCAATC	1333	0.337222480824108	No Hit
00071470004AACTTOTACAGOTATATATCONTICATIONADO 13124 407932314964 No Hit 00070470004AACGATAAACTCONTOGATAGOCAATATAT 1118 0.3970244847317897 No Hit 000400047004AAACTCONADATATAT 1118 0.397024484513189 No Hit AAGCCARMAAACTCONADATATAT 1118 0.387024484513189 No Hit AAGCCARMAAACTCONADATATATATATATA 113 0.386124649520364 No Hit AAACTCONADATATATATATATATATATATA 113 0.386124094203264 No Hit AAACTCONADATATATATATATATATATATATATATATATATATATA	CGATAAAACTCTGCAGGTTGGATACGCCAATCATTTTAT	1304	0.32988605776041774	No Hit
00071470004AACTTOTACAGOTATATATCONTICATIONADO 13124 407932314964 No Hit 00070470004AACGATAAACTCONTOGATAGOCAATATAT 1118 0.3970244847317897 No Hit 000400047004AAACTCONADATATAT 1118 0.397024484513189 No Hit AAGCCARMAAACTCONADATATAT 1118 0.387024484513189 No Hit AAGCCARMAAACTCONADATATATATATATA 113 0.386124649520364 No Hit AAACTCONADATATATATATATATATATATA 113 0.386124094203264 No Hit AAACTCONADATATATATATATATATATATATATATATATATATATA	TAAAAATGATTGGCGTATCCAACCTGCAGAGTTTTATCGC	1277	0.32305559490801644	No Hit
BOANGORIAMANCTERCONSTRUCTATION 1118 0.39202448971797 No. BIL ANGORIAMANCTERCONSTRUCTATION CONSTRUCT 113 0.38715460013131 No. BIL ANGORIAMANCTERCONSTRUCTATION CONSTRUCT 113 0.3871546001370 No. BIL AMARTERGONSTRUGATION CANTERTIATION III 0.3871546001370 No. BIL AMARTERGONSTRUGATION CANTERTIATION III 0.38715460137074 No. BIL ANGORIAMANCTERGONSTRUGATION CANTERTIATION IIII 0.38716460137074 No. BIL ANGORIAMANCTERGONSTRUGATION CONTENTIATION IIIII 0.3871640137074 No. BIL ANGORIAMANCTERGONSTRUGATION CONTENTIATION IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII				
BOANGORIAMANCTERCONSTRUCTATION 1118 0.39202448971797 No. BIL ANGORIAMANCTERCONSTRUCTATION CONSTRUCT 113 0.38715460013131 No. BIL ANGORIAMANCTERCONSTRUCTATION CONSTRUCT 113 0.3871546001370 No. BIL AMARTERGONSTRUGATION CANTERTIATION III 0.3871546001370 No. BIL AMARTERGONSTRUGATION CANTERTIATION III 0.38715460137074 No. BIL ANGORIAMANCTERGONSTRUGATION CANTERTIATION IIII 0.38716460137074 No. BIL ANGORIAMANCTERGONSTRUGATION CONTENTIATION IIIII 0.3871640137074 No. BIL ANGORIAMANCTERGONSTRUGATION CONTENTIATION IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	TGCGTCATGGAAGCGATAAAACTCTGCAGGTTGGATACGC	1233	0.3119244702596588	No Hit
ARCT000000TT0000TX000CACMACTTTFXT0C00M.00 113 0.34812443945320954 No. Hit AMARCT000000TT0000TX000CACMACTTTTTC00M 1131 0.34812449945320954 No. Hit AMARCT000000TT0000TX000CACMACTTTTTC00M 1131 0.34812449945320954 No. Hit AGGMTX00000TT0000TX000CACMACTTTTTC00M 1131 0.3481244994532045 No. Hit AGGMTX00000TT0000TX000CACMACTTTTTTC00 110 0.34812483974531298 No. Hit AGGMTX00000TT0000TX000CACMACTTTTTTC00 100 0.349397145412985 No. Hit AGGMTX000CANTACOCCAATCACTTTTTATC01M 100 0.349397145412985 No. Hit TTT0000TX000CACMACTACATCTTTATC01M 100 0.339317397145412985 No. Hit TT0000TX000CACMATAATCT000AGGTT00ATXACCC 0.339317397145412985 No. Hit GAGMACMATAATACT000AGGTT00ATXACCC 0.33931739714541298 No. Hit GAGMACMATAATACT000AGGTT00ATXACCC 0.33931739714541298 No. Hit GAGMACMATAATACT000AGGTT00ATXACCC 0.339314973978541 No. Hit GAGMACMATAATACT000AGGTT00ATXACCC 0.32934498323208 No. Hit GAGMACMATAATACT000AGGTT00ATXACCC 0.13939344983252181 No. Hit	GGAAGCGATAAAACTCTGCAGGTTGGATACGCCAATCATT	1182	0.2990224848717897	No Hit
AMACTECTOCADAGTOCATACATETTATATCOMA 11.11 0.3815429372116 No. III. AMACTECTOCADAGTOCATACATETTATATCOMA 0.381546837582119 No. III. AMACTECTOCADAGTOCATACATETTATATCOMA 0.381546837582119 No. III. ARCECTOCADAGTOCATACATETTATATCOMA 0.3815468375821197 No. III. ARCECTOCADAGTOCATACAGCCAATCATETTATACOMAC 0.3815468971313074 No. III. ARCECTOCADAGTOCATACACCCAATCATETTATACOMAC 0.393754544147958 No. III. CATADAATCETOCADAGTOCATACACCTATATTATACOMAC 0.393754544147958 No. III. CATADAATCETOCADAGTOGATACOCCAATCAT 0.393751434177458 No. III. TOCADACGATAMAATCETOCADAGTOGATACOCCAATCAT 0.393751434177746 No. III. COCATAMAAATCETOCADAGTOGATACOCCAATCATT 88 0.324464325418 No. III. COCATAMAAATCETOCADAGTOGATACOCCAATCATT 88 0.31451358018155 No. III. COCATAMAAATCETOCADAGTOGATACOCCAATCATT 0.393714644381655 No. III. COCATAMAAATCETOCADAGTOGATACOCCAATCAT 0.393714644381615 No. III. COCATAMAAATCETOCADAGTOGATACOCCAATCAT 0.3937146446381613 No. III. COCATAMAAATCETOCADAGTOCATACOCCAATAT 0.393714644648161	AAGCGATAAAACTCTGCAGGTTGGATACGCCAATCATTTT	1136	0.2873854000121431	No Hit
AMACTECTOCADAGTOCATACATETTATATCOMA 11.11 0.3815429372116 No. III. AMACTECTOCADAGTOCATACATETTATATCOMA 0.381546837582119 No. III. AMACTECTOCADAGTOCATACATETTATATCOMA 0.381546837582119 No. III. ARCECTOCADAGTOCATACATETTATATCOMA 0.3815468375821197 No. III. ARCECTOCADAGTOCATACAGCCAATCATETTATACOMAC 0.3815468971313074 No. III. ARCECTOCADAGTOCATACACCCAATCATETTATACOMAC 0.393754544147958 No. III. CATADAATCETOCADAGTOCATACACCTATATTATACOMAC 0.393754544147958 No. III. CATADAATCETOCADAGTOGATACOCCAATCAT 0.393751434177458 No. III. TOCADACGATAMAATCETOCADAGTOGATACOCCAATCAT 0.393751434177746 No. III. COCATAMAAATCETOCADAGTOGATACOCCAATCATT 88 0.324464325418 No. III. COCATAMAAATCETOCADAGTOGATACOCCAATCATT 88 0.31451358018155 No. III. COCATAMAAATCETOCADAGTOGATACOCCAATCATT 0.393714644381655 No. III. COCATAMAAATCETOCADAGTOGATACOCCAATCAT 0.393714644381615 No. III. COCATAMAAATCETOCADAGTOGATACOCCAATCAT 0.3937146446381613 No. III. COCATAMAAATCETOCADAGTOCATACOCCAATAT 0.393714644648161	ACTCTGCAGGTTGGATACGCCAATCATTTTTATCGAAGCG	1133	0.28662645969520956	No Hit
AAAACCTOCOGADOTOGATCOCACTCATTTATCOGA 1129 0.3854637524514 80 Hic ACOGNTAGACCTOCACTCATTTATCOGA 111 0.38166457351375 No Hic ATAMACTOCOGNTOGATACOCCAATCATTTATCOGA 111 0.38166457351375 No Hic ACCTTOCAGOTTOGATACOCCAATCATTTATCOGAL 103 0.3739774641385 No Hic ACCTTOCAGOTTOGATACOCCAATCATTTATCOGAL 103 0.3739774641385 No Hic ACCTTOCAGOTTOGATACOCCAATCATTTATCOGAL 0.333917978641392 No Hic ACCTTOCAGOTTOGATACOCCAATCATTTAT 0.333917978641392 No Hic ACCATTOCAGOTTOGATACOCCAATCATTTAT 0.333917978641329 No Hic ACACCTATAGATCOCCAATCATTTAT 0.323816379804984 No Hic COTATAGATCAACCTOCAATCATTTAT 0.32484439321329 No Hic COTATAGATCAACCTOCAATCATTTAT 0.3393147978641339 No Hic COTATAGATCAACCTOCAATCATTTAT 0.3393402825218 No Hic COTATAGATCAACCTOCAATCATTTAT 0.33934028252183 No Hic TOCOCACTOCAAAATTTATACOCTOCAATCAT 0.34				
ACCUSTANALCTCCOCMATCASTANCCCCANTCASTT 111 0.381666973781097 No. Bit ATAMACTCTCCCASTANCCCCANTCASTTTATCCG 111 0.38166697371097 No. Bit ATAMACTCTCCCASTANCCCCANTCASTTTATCCG 111 0.38166697371097 No. Bit CTCCCASTANCACCCANTCASTTTATCCAMACC 103 0.27397748412985 No. Bit CTCCCASTANTTATCCCCATCASTTTATCCAMACC 103 0.27397748412985 No. Bit CTCCCASTANANCTCTCCCAGCAGAMACTTAACK 0.33957160047774 No. Bit CTCCASTANANCTCTCCCASTANCTTTATCCAST 0.33957160047774 No. Bit CCCASTANANCTCTCCCAGATCASTTTATCA 0.33957160047774 No. Bit CCCASTANANCTCTCCCAGATCASTTTATCCAST 0.3395716004781155 No. Bit CCCASTANANCTCTCCCAGATCASTTTATCCAST 0.339571604781155 No. Bit CCCASTANANCTCTCCCAGATCASTTTATCCASTTTAT 0.3395716047851155 No. Bit CCCASTANANCTCTCCCAGATCASTTATCCCASTTTAT 0.3395716047655113 No. Bit CCCASTANANCTCTCCCAGATCASTTATCCCASTTAT 0.3395716047695113 No. Bit CCCASTANANCTCTCCAGATCASCASTTAT 0.3395716047695113 No. Bit CCCASTANANCTCCCACASCASTTATATCCASCASTTAST 0.3395716047697361				
AACTCTGCAGGTTGGATAGGCCARCATTTTATCGAGG 193 0.739745412965 No Bit CTGCALMODITTTATGCTCCARLOCALMONTTATACCARCE 105 0.26849011454932 No Bit TECCONTGGATAGGCATTAGGCARGATTGATA 105 0.239721604313402 No Bit TECCONTGGATAGGCATAAGCCTCGAGGGTTGGATAFGCACCARCATT 0.239721604371778 No Bit TEGGATGGATAGGCATACATTTTTTTGGA 0.239721604313402 No Bit CAMACTCTCGAGTTGGATAGCCCARCATTTTTTTGGA 0.239721604318155 No Bit CAMACTCTGCAGGTGGATAGCCCARCATTTTTTTGGA 0.239721604318155 No Bit COGATAMAACTCTGCAGATTGGATAGCCCARCATT 0.13921640245218 No Bit COGATAGAAGCCAGGAGGTTGGATAGCCCARCATT 0.13921640245218 No Bit TEGGCATGGAGGATGGAGGATGGATGGATGGAT 74 0.13921640245218 No Bit TEGGCAGTAGGAGGAGGATAGACCCARCATTTAT 0.13921640245218 No Bit CCAACTGGAGAGTTTATTATGCGCACGATGGATAGC 0.13921640245218 No Bit TEGGCAGTAGGAGGATAGACCCARACGATTAGC 0.139216402462180 No Bit CCACGGAGGAGGATAGACCCARAGGTGGAGGTGGATGAT 0.1393768037603760 No Bit CCTGCAGGTGGAGGAGGATAGACCCAGGAGGTGGAGGTGGAT 0.1394510428277 No Bi				
AACTCTGCAGGTTGGATAGGCCARCATTTTATCGAGG 193 0.739745412965 No Bit CTGCALMODITTTATGCTCCARLOCALMONTTATACCARCE 105 0.26849011454932 No Bit TECCONTGGATAGGCATTAGGCARGATTGATA 105 0.239721604313402 No Bit TECCONTGGATAGGCATAAGCCTCGAGGGTTGGATAFGCACCARCATT 0.239721604371778 No Bit TEGGATGGATAGGCATACATTTTTTTGGA 0.239721604313402 No Bit CAMACTCTCGAGTTGGATAGCCCARCATTTTTTTGGA 0.239721604318155 No Bit CAMACTCTGCAGGTGGATAGCCCARCATTTTTTTGGA 0.239721604318155 No Bit COGATAMAACTCTGCAGATTGGATAGCCCARCATT 0.13921640245218 No Bit COGATAGAAGCCAGGAGGTTGGATAGCCCARCATT 0.13921640245218 No Bit TEGGCATGGAGGATGGAGGATGGATGGATGGAT 74 0.13921640245218 No Bit TEGGCAGTAGGAGGAGGATAGACCCARCATTTAT 0.13921640245218 No Bit CCAACTGGAGAGTTTATTATGCGCACGATGGATAGC 0.13921640245218 No Bit TEGGCAGTAGGAGGATAGACCCARACGATTAGC 0.139216402462180 No Bit CCACGGAGGAGGATAGACCCARAGGTGGAGGTGGATGAT 0.1393768037603760 No Bit CCTGCAGGTGGAGGAGGATAGACCCAGGAGGTGGAGGTGGAT 0.1394510428277 No Bi				
CTOCADACTITATEOCTTCONDUCCADACTITATACC 105 0.368931043932 No. Bit TECENCENTONICATINANCTCONDUCTATIONITY NY 0.339331017933402 No. Bit TECENCENTONICATINANCTCOCADATICAT 146 0.33933101795349754 No. Bit TEAMACTCTCOCADGTTGGATACCCCANTCATTTATACCE 0.230473101804177768 No. Bit GOUTANALCCCTOCADGTTGGATACCCCANTCATTTE 0.23047310190329 No. Bit GOUTANALCCCTOCADGTTGGATACCCCANTCATT 0.230448930439363 No. Bit GOUTANALCCCTOCADGTTGGATACCCCANTCATT 0.130531002300320 No. Bit GOUTANALCCCCTOCADGTTGGATACCCCCANTCATT 0.1393314023823103 No. Bit CONTANALCCCCCCOCADGTTGGATACCCCCANTCATT 0.1393304028232103 No. Bit TEGEOGTATAALCCCCCCCANTCATTTATA 0.1393304028232103 No. Bit TECONCCCCCANAGTTTATACCCCCCANTCAT 0.1393304028252103 No. Bit CCANCCCCANAGTTTATACCCCCCANAGTTGAT 0.138231022810304463802 No. Bit CCANCCCCANAGTTTATACCCCCCANAGTTGATA 0.138231022810304 No. Bit CCANCCCANAGTTATACCCCANAGTTGATACCCUNAT 10.13135022810307 No. Bit CCANCCCANAGTTATACCCUNATACT 0.131335022810307 No. Bit COTTAGAGGA				
TTCTCCCTCCTCATGAMACCTCTCCAACGCAATCAT 0.3393121604531462 No. Hit TGGAMGCGATMAMACTCTCGCAACGCAATCAT 946 0.33931179336854 No. Hit TGGAMGCGATMAMACTCTCGCAACGCAATCAT 946 0.33931179336854 No. Hit GAMACTCTCCCAAGTGGATAGCCCAATCAT 948 0.239419179336854 No. Hit GAMACTCTCCCAAGTGGATAGCCCAATCAT 948 0.2444633381239 No. Hit GCAATGAAAACTCTCCCAAGTGGATAGCCCAATCAT 948 0.2394198381615 No. Hit GCGATGAAAACTCTCCCAAGTGTGAAACTCCCAATCAT 0.193426432381615 No. Hit GCGATGAAAACTCTCCCAAGTGTGAATCCCCAATCAT 0.193247040523118 No. Hit GCGATGAAAACTCTCCCAAGTGTGAATCAC 1.9393463263218 No. Hit TCCCACTGAAGAGTTTAATCCCCAATCATGTGAATA 1.9393463253218 No. Hit TCCCACTGAAGAGTTTAACCCAATAACTCTCCAAGTGGAAGT 1.93834734053128 No. Hit TCCCACATGGAAGTATAAACTCTCCAAGCGAAGTGTGAATA 1.9314934934936128 No. Hit TCCCACATGGAAGTATAAACTCCCACACGCAAGTGGAAGTGTAACC 1.9314934934930191 Hitmina Paired End PCR Primer 2 (964 over 2 CCCTATAAGGCAATAAACTCTCCACGCAGGAAGTGTAAGCCCC 1.934593418028279 No. Hit CCCTATAAGGCAATAAACTCTCC				
TOCAMACCARCHARANCECCOCARCAT 946 0.339113793946934 No. Bit TAMACCTCOCAGTTGGATAGOCGARCATTITATOGA 912 0.33971365437756 No. Bit GANCGARTAACCTTOCAGTTGGATAGOCCARCATT 94 0.32444533131309 No. Bit GOUTANGACTGARTGGATGGATAGOCCARCATT 94 0.34444533113109 No. Bit GOUTANGACTGARGACTOCAGTTGGATAGOCCARCATT 0.1348493634311309 No. Bit CONTANAACTCOCAGTTGGATAGOCCARCATTITA 0.134931463282348 No. Bit TOGOCATAGAACCTACACCARCATTITA 0.134931463282348 No. Bit TOGOCATAGAACCTACACCAGANGATA 0.134931463282348 No. Bit TCCANCTOCAGATTTATACOCTCACATACOCAMUT 14 0.138314038253183 No. Bit TCCANCTOCAGATTATACCCCACANTATA 15 0.1382130944585012 No. Bit CCANCTOCAGATATACCCCAANTATA 15 0.1382130944585012 No. Bit CCANCTOCAGATATACCCCATACACTCAGACOCAMUT 14 0.13851403827895012 No. Bit CCANCTOCAGATATACCCCATACACTACAGACAGANGTTGATACC 10.138140389399856 No. Bit CCANCTAGATATACCCATACATACAGAANTATACCACCANA 13 0.1515309876110011 CCANCTAGATATATACCCATACANANANACTT				
TANAGCTOCOCONTGGATAGOCCANCATTITATCGA 912 0.230717865477766 No Bit GAMCGENAMAGCTOCOCONTGGATAGOCCANCATTI 88 0.23644989581815 No Bit GAMCGENAMAGCTOCOCONTGGATAGOCCANCATTI 88 0.23644995081815 No Bit GCGATAGAMAGCTOCOCONTGGATAGOCCANCATTITAT 0.33644995081815 No Bit GCGATAGAMACTOCOCONTGGATAGOCCANCATTITAT 0.3954995082128 No Bit GCGATAGAMACTOCOCONTGGATAGOCCANCATTITAT 0.19821970840581103 No Bit GTCOTOCOTAGGAGACGATAGOCCANATATTTAT 0.19821971985950212 No Bit CCMACTOGANGOCATAGAACCTOCANCONTOGATAGOCCANAT 0.19821971985950212 No Bit CCMACTOGANAGATAGOCCANAACTOTOCANCATAGOCCANAT 0.19821971985950212 No Bit CCMOCTOGANGACATAAACTOCTOCANTOGATAGOCCAN 0.198219719859950212 No Bit CCMOCTOGANGACCANAACTOCTOCANTOGATAGOCCAN 0.198219719859950212 No Bit CCMOCTOCANGACCANTAAACTOCTOCANTOGATOCANT 0.198219719859950212 No Bit CCMOCTAGAGACANTAAACTOCTOCANTOGATAGOCCAN 0.198219719959950212 No Bit CCMOCTAGAGACANTAAACTOCTOCANTOTTOGATACCQ 0.198219719985980212 No Bit CCMOCTAGAGACANTAANACTOCTOCANTUTTATTAGOCANTOGATACCQ				
GAMCGATABAACTCTGCKGGTGGATACGCCARTCHT 88 0.224446333812369 NO BIE GGTGTAGGGAGGUAGAGUATABACTGTGCKGGGGATGGGATAGGCCARTTHT 88 0.124464333812369 NO BIE GGTGTAGGGAGGUAGGUAGGATGGATGGCCARTCHTT 88 0.158858323823818 NO BIE TGGGGTAGGAGGUAGGUAGGUAGGUAGATGTTGT 0.15885843283818 NO BIE TGGGGTAGGUAGGUAGGUAGGUAGATGTTGT 0.158258423818 NO BIE TCCAACGGCAAGUATGGUATAGGCAGTGATGT 0.15825042385218 NO BIE TCCAACGGCAAGUATGGUATAGGCCAAGUTGT 0.1582505855021 NO BIE TCCAACGGCAAGUATGGUAGGUAGGUAGGUAGGUAGGUAGGUAGGUAGG				
GOTGATOGUARGUARTANALCTFOCOLOGITOGUARGOCE 0.301448998.04384.015 No. Hit GOGATNAMACTCTOCANGOCTOGUARGOCE 0.1985.983.02302.012 No. Hit GOTGATNAMACTCTOCANGOCTATURATTATA 0.1985.983.02302.012 No. Hit GOTGATOGUARGOCANAGACTTATURGOCTACUATURAT 0.1985.983.02302.012 No. Hit CTCUTCOCTATUGAGOCTATURGOCTACUAGUAGUARGA 0.1982.01244448466 No. Hit CCANCCTOCUARGUARGUARGUARGUARGUARGUARGUARGUARGUARG				
OCIMINAMACTCTOCADGTOCATACACTOCATACA 715 0.1985833823093612 No Bit TODOCTATACADACTOCAMACATOCATACA 714 0.1985833823093612 No Bit TODOCTATACADACCOMATACATOCATACA 712 0.1983840282318 No Bit TCCMACTOCAMACTTATACOCCAMATOT 0.19838402825218 No Bit TCCMACTOCAMACTTATACOCCAMATOT 0.19828403825218 No Bit TCCMACTOCAMACTTATACOCCAMATOT 0.19828403825218 No Bit TCATOGNAMATTTATACOCTACATAGTOCAMATT 0.198284038445806 No Bit TCATOCANCATANACTCTOCACOCAMATT 0.188211702538724 No Bit COTOCANACTACACACATAACCTACACACAMATTATAC 0.1882107825389454 No Bit COTOCANGTATAANCTCOCACACOCAMATTATACCAMAT 0.158510821089454 No Bit COTOCAAGTATAACCCACACACACACAGTACACCAMATTATACCAMATTATACAMACOCTAC 0.15851082108108 Ni Bit COTOCAAGTACACACATACACACACAGTAGACTACACACAGTATATATA				
TODGOTATICAACCTGCAMAGTTTTATCGCATCGCATGA 744 0.1983344028252818 No Hit CTTCTGCGTCATGAMAGCGATAAACTCTGCAGTTGCATGA 0.1983344028252818 No Hit CTCCGCGTCATGAMAGCGATAAACTCTGCAGTTGCGATGA 0.192737684581163 No Hit CCAACCTGCAAAGAGTTATACGCGCTACAATT 0.19281374468896 No Hit CCAACCTGCAAAAAACTGCCAATGATGACCAAAATT 0.19381378458956212 No Hit CTGCGACGTGAAAAAACTGCCCAATGACGAAAAATTAAC 0.18581372357875 No Hit CCTGCAAGTGGAAGTCGCAATGACGCAAAAATTAAC 0.15851892623989656 No Hit CCTGCAAGTGGAATGCCCAAAACGCGAAAAATTAAC 0.158518926219097 No Hit CCTGCAAGTGGAATGCCCAAAACGTGCAAGTGGTAGC 50 0.151353091218077 No Hit CCTGCAAGTGGAACGCAAAAACTGCAAGTGGTAAGCGGTTAGC 50 0.13458913179955 No Hit CCTGCAAGGTGGAATGCCGAAGTGGAATGAAGGGGGTGGAATGCGCAATGAAGTGGTAAGTGGGAAGGAA				
CTCTOCOCCATORUMACCETTORADECTOCAMONTORY 0.192779840551103 No. Bit TCCAACCTOCAMAGUTTARCOCCATORADACT 752 0.193241034445866 No. Bit TCCAACCTOCAMAGUTTARCOCCATORADACT 752 0.193241034445866 No. Bit TCATORADACTATACOCCATAGUCAGAMACT 752 0.193211702535724 No. Bit TCATOGAACMAGUTTARCOCCATACUCAGAMACT 0.1581217192539724 No. Bit CCTOCAAGUTAGUCATAAACCTOCACAGTORADACCOTACA 0.158153021398955 No. Bit CCTOCAAGUTTARCOCCATACUCAGUAGUTTARCACCAA 0.158550213998955 No. Bit CCTOCAAGUTTARCOCCATACUCAGUAGUTTARCACCAA 0.158550213981098 TLUMIAN Price End PCR Primer 2 (964 over 2) COTTACAGUAGUTARCOCATACUTTTATACOGUMOCOTTACC 50 0.13984510183157965 No. Bit CTCOCACUTACUAGUTARCOCATACUTTTATACOGUMOCOTTACCAGUT 52 0.13984510183157965 No. Bit CTOCACUTACUTAGUTARCOCATACUTTTATACOGUMOCOTTACUTORATOR 1.3384531024898 No. Bit CTOCACUTAGUTAGUTAGUTAGUACOUCO CTOCACUTAGUTAGUTARCOCATACUTTTATACUAGUCCOUCO 0.13984531024898 NO. Bit CTOCACUTAGUTAGUTAGUTAGUACOUCOUCOUCAUCUTTATATACUTAGUAGUCCOUCOUCUTTTATACUTAGUAGUCCOUCUTTURATACUTAGUAGUCCOUCUTTURATACUUCUTUTUTUTURAGUUCUUCUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU				
TCOMACTOCAMANTTARACCONCAMANT 742 0.198341034445866 No. Bit COMACCEGNAMENTTARACCONCAMANT 744 0.18821135559212 No. Bit TCANCINAMENTARACCTOCAMOTTOGATACOCAMANT 744 0.18821135559212 No. Bit TCTAOGAMCOGATAMAACTCOCAMOTTOGATACOCAMANT 744 0.18821135559212 No. Bit CONTANDAMACCONTANAMACTCOCAMOTTOGATACOCAMANT 744 0.15821377025389726 No. Bit CONTANDAMACTCOCAMANTOCAMANTARA 43 0.155076807607164 No. Bit CONTANDAMACTCOCAMATOCOMALINITARA 43 0.15507680707164 No. Bit CONTANDAMACTACOCAMATTARA 43 0.15307680707164 No. Bit CONTANDAMCTACACATTATATACAMOCOMOTTOGATAC 0.1479391368077 No. Bit CONTANDACTACCATTATATACOCAMOTOGATACATTAT 0.136456141507184 No. Bit CONTANDACTACACATTATATACOCAMOTOGATATOCATICATACOMOCOMO 50 0.1364561412628498 No. Bit CONTANDACTACCATATATATACOMACOMOTOGATOCA 1.148456181373845 No. Bit CONTANDACTACACATTATATACOMACOMOTOGATOCA 0.138475446591342 No. Bit				
CCAACCOGLAMACTITATCCCTCCATGLOCCAALANT 744 0.188119189595212 No Hit TCATGGLAMACCTCCCAGGGTGGATAGCOCLAM 744 0.188219196595212 No Hit TCATGGLAMAGCGLTAALACCTCCCAGGTGGATAC 453 0.18821936454 No Hit CCTCGLAMAGTTATATCCCCTCCAGGTGGLAMAGTTAAC 453 0.155520921510078 111mina Paired End PCR Primer 2 (96% over 2 TCTCCAGGTGGLAMACCGLAMACTTAACCGCTCGCG 595 0.155535082150078 111mina Paired End PCR Primer 2 (96% over 2 TCTCCAGGTGGLAMACCGLAMACTTAACCGCCGGGGLAMAGTTAACCGCCAGGTGGLAMACGGLAMAGTTAATGCCGGAGGGGATTTATTCCGAAGCGG 595 0.13458911679855 No Hit CTCTCCAGGTGGLAMACGGLAMACTTATCGLAMAGCGC 522 0.13458911679855 No Hit CTCCCAGGTGGLAMACGGLAMACTTATCGLAGGCGG 515 0.13458911628698 No Hit CTCCCAGGTGGLAMACGGCAMACGTCTCGLAGGTGGLAMAGTGGGLAMAGTGGLAMAGTGGLAMAGTGGLAMAGTGGLAMAG				
TOATGUARGUERAMANCTGUARGUTGUATEGOCIATE TOATGUARGUERAMANCTGUARGUTGUATEGOCIATE CETOGURAMONGUERAMANCTGUARGUTGUARGUERA CETOGURAMONTTARCGURAMUGGUTGUARGUERA COTTAGURAMONTTARCGURAMUGGURAMUGTARGA 617 0.151350828180578 JILMEIA Prime 2 (564 over 2 TORUGAUTGUARGUARGUARGUTGUARGUCGUTTUR 0.134536180578 Mo ILL COTTAGURAGUTARGUARGUTGUARGUTGUARGUERA 0.13454518180578 Mo ILL COTGURAGUTGURAMUGGURAMUTTURGURAGUERA 0.1345451818278 Mo ILL COTGURAGUTGURAMUGGURAMUTTURGURAGUERA 0.134584182824898 Mo ILL COTGURAGUTARGURARGUARGUERTURGURGUERA 0.1275459138264852 Mo ILL CONCORDIGURAMUGGURANUTURGURGUERAGUERAGUERA 0.1275459138264852 Mo ILL CONCORDIGURANUTURGURAGUERAGUERAGUERAGUERA				
TCTGCGTCATGGAGCGGATAAAACTCTGCAGGGTTGGATAC 427 0.15861852623998656 No Hit CCTGCAGAGTTTTATGGCGTCGAGCGGAAAGTTAAAC 0.15553082169078111mina Paired End PCR Primer 2 (96% over 2) CCTGCAGGTGGATCGGAAAGTTAAACCGGGGGTTGGGTGGATGGA				
CCTUCARAGENTTINCCCTCCCARACCTARACCARAGE 413 0.1557968474097366 No Hit COGTADACAGAANTATANCCCTARACCARACGTARAC 539 0.1515350822810578 Tilumine Paired End PCR Primer 2 (964 over 2 TOCAGATTGARACCARACARTTATTGARCGARAGEO 599 0.1179391581802379 No Hit CCTTARAGENACCARTANTATCGARGOGO 500 0.13458415282498 No Hit CTCCCARGTGARGACCARTANTTGARCGARC 515 0.13458415282498 No Hit CTCCARGTGARGACARACARTTGARCGARC 515 0.132947544051342 No Hit CTCCARGTGARGACCARACTTGARCTGARGACCO 515 0.1327847540564823 No Hit				
COSTCAGCAGGANCOCCOAGATCOGAAGAGGGGTCAGC 599 0.15135308228108078 111umina Faired End FCR Frimer 2 (96% over 2 retrocAGGTGGAAFAGGCGAACATTFTATCGCAAGGGG 585 0.191933161802079 No Hit COSTCAAGCTACATTATGGCGGGGGGGTTTTTTT 52 0.13465811573965 No Hit 0.110 COSTCAAGCTACACATTTTATGGCGAGGGC 512 0.13465811573965 No Hit 0.110 COSTCAAGCTACACCATTTTATGGCGGGGGGG 515 0.134658116228698 No Hit 0.1026845116228698 COSCGATGGAAGGCGCACACATTTTTATGGCAGGGG 510 0.1302847346691312 No Hit 0.110				
TCTORADATIONATARGOCANTATITTATICAMACON 585 0.137933161802877 No. Bit CGCTTAMAGCIACCAGTACGOGOGOGITTTTTTT 552 0.1384501831573945 No. Bit CTOCAGATAGGAGACAGTATATAGGICGOGOGOTTTTTTT 552 0.138458162895 No. Bit CTOCAGTAGGAGACAGTATATAGGICGOGOGOTTGTTAGGICGOGOGOTTGGATAGGICGOGOGOTTGGATAGGICGOGOGOTTGGATAGGICGOGOGOTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTAGGICTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTAGGICTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTTGGATAGGICGOGOSOTTGGATAGGICGUCGUCGUCGUCGUCGUCGUCGUCGUCGUCGUCGUCGUC				
CGCTRAMACTACCAGTTATATGGCGGGGGGTTTTTTT 552 0.13944501831535965 No Hit CTCTCGCGGTGGGATAGCCAARCATTTTATGCGAGCGC 532 0.134585145228598 No Hit CTCGGGTGATGGAARGCGATAAAACTGGCGGTGGGTGGG 515 0.13028475440651342 No Hit CTCGGGTGGTGGGAAGCGATTTATTGATGGAGCGGG 55 0.1275485313064852 No Hit				
CTCTGCMGGTTGGMTACGCCMATCATTTTATGAMGGCC 512 0.1345854162028698 No Hit CTGCGTAGGMAGGATAMAACCGTGGGATGGTAGGATGG 0.132847344653182 No Hit CGCAGGTGGGATGGATTGTTGGAGGGCGC 550 0.1273459339464823 No Hit				
CTGCGTCATGGIAGCGATAAAAACTCTGCAGGTTGGATACC 515 0.13028475440691342 No Hit CTGCAGGTTGGATACCCCAATCATTTTATCGAAGCGCGC 505 0.12775495335046952 No Hit				
CTGCAGGTTGGATACGCCAATCATTTTTATCGAAGCGCGC 505 0.12775495335046852 No Hit				
	CTGCGTCATGGAAGCGATAAAACTCTGCAGGTTGGATACG	515	0.13028475440691342	No Hit
GCTTABAGCTACCAGTTATATGGCTGGGGGGGTTTTTTTTTG 411 0.10397482341988626 No Hit				
	CTGCAGGTTGGATACGCCAATCATTTTTATCGAAGCGCGC	505	0.12775495335046852	No Hit

Adapter Content

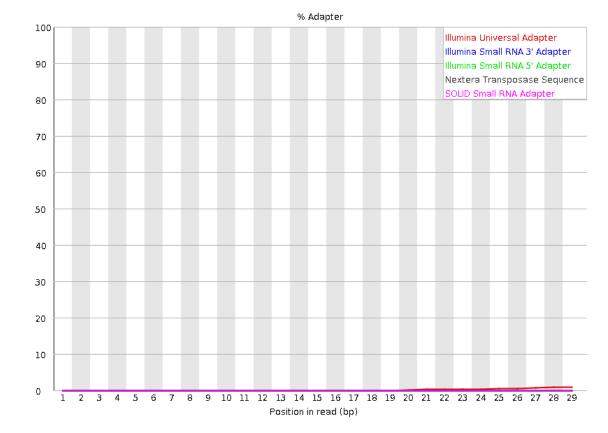
Good Quality

Adapter Content



Bad Quality

Adapter Content



Run FastQC

- 1. Go to the RNA-seq data directory
- 2. Make a directory to put the FastQC reports into, fastqc
- 3. Run fastqc on the samples
- for i in *.fastq.gz; do fastqc \$i -o fastqc/;
 done

Trim Bad Quality Sequences

 Trimming removes sequencing adapters, bad quality sequences, and/or other biased sequence information

- Trimming removes sequencing adapters, bad quality sequences, and/or other biased sequence information
- Why is that important?
 - Helps prevent incorrect base calls by removing poor quality information
 - Increases speed and accuracy of alignment by removing artificial sequences and low quality sequences

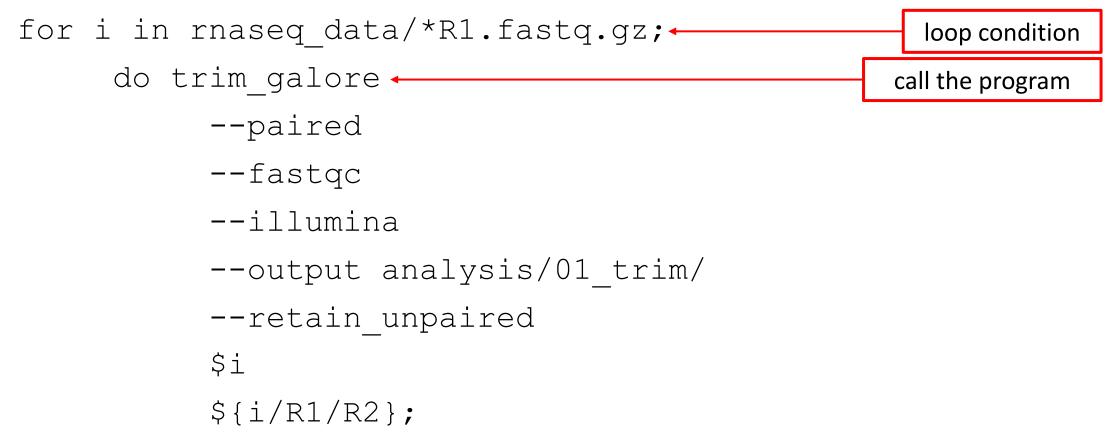
- Trimming removes sequencing adapters, bad quality sequences, and/or other biased sequence information
- Why is that important?
 - Helps prevent incorrect base calls by removing poor quality information
 - Increases speed and accuracy of alignment by removing artificial sequences and low quality sequences

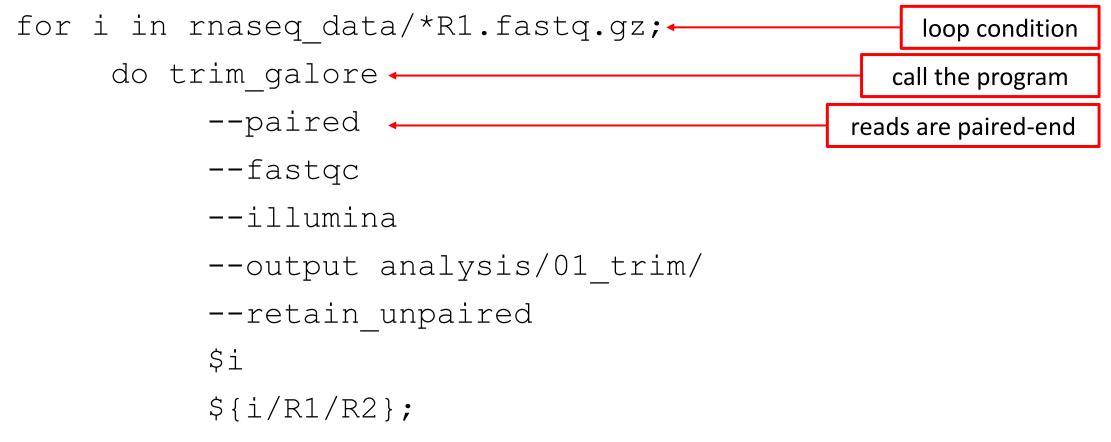
- Trimming does two complementary things:
 - Removes any sequence information that comes from library preparation or sequencing
 - 2. Removes low quality bases / low quality reads

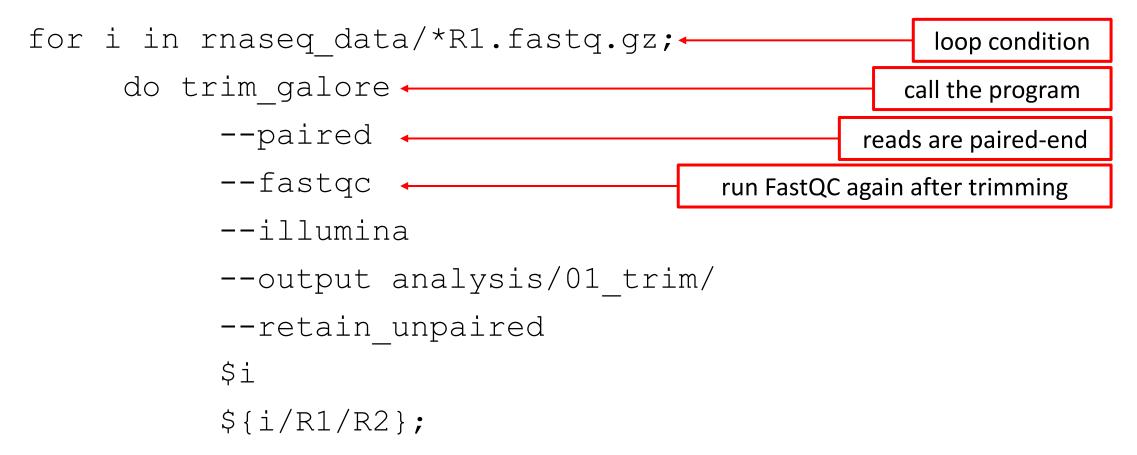
- 1. Go back up to the rnaseq directory
- 2. Make a folder to put the analysis results in, analysis
- 3. Make a folder inside the analysis folder to put the trimmed reads in, analysis/01_trim

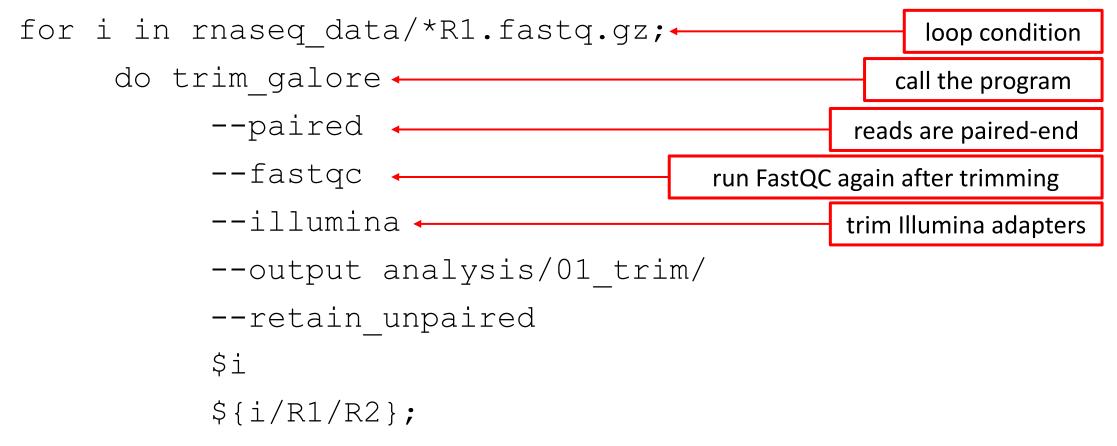
for i in rnaseq data/*R1.fastq.gz; do trim galore --paired --fastqc --illumina --output analysis/01 trim/ --retain unpaired \$i \${i/R1/R2};

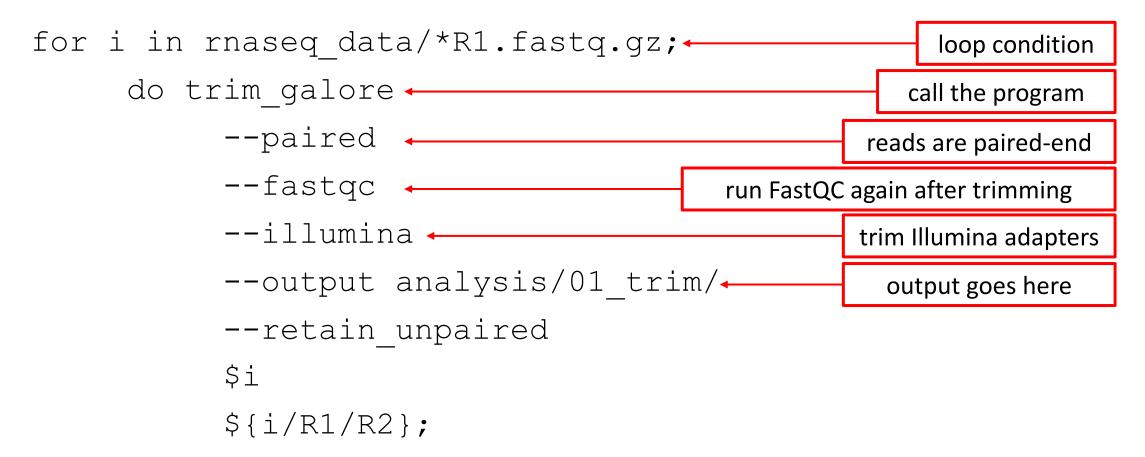
for i in rnaseq data/*R1.fastq.gz; loop condition do trim galore --paired --fastqc --illumina --output analysis/01 trim/ --retain unpaired \$i \${i/R1/R2};

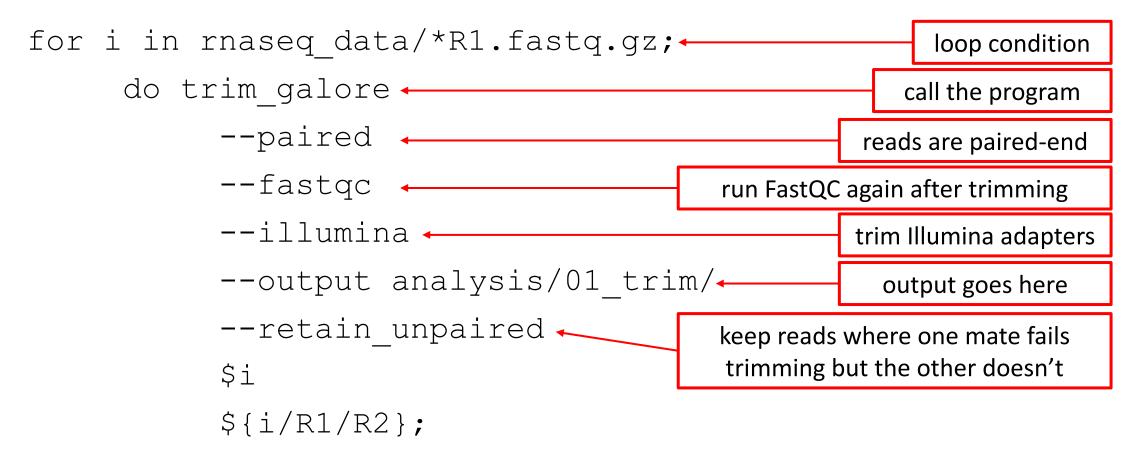


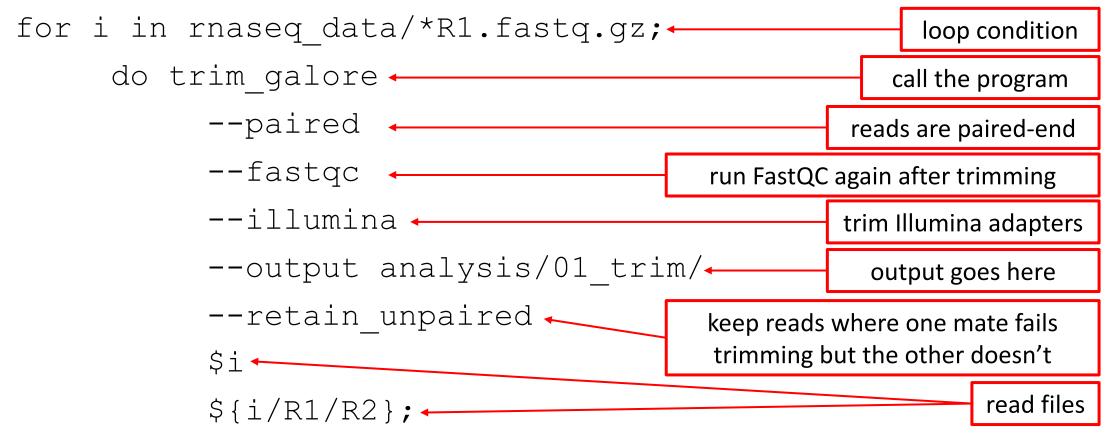


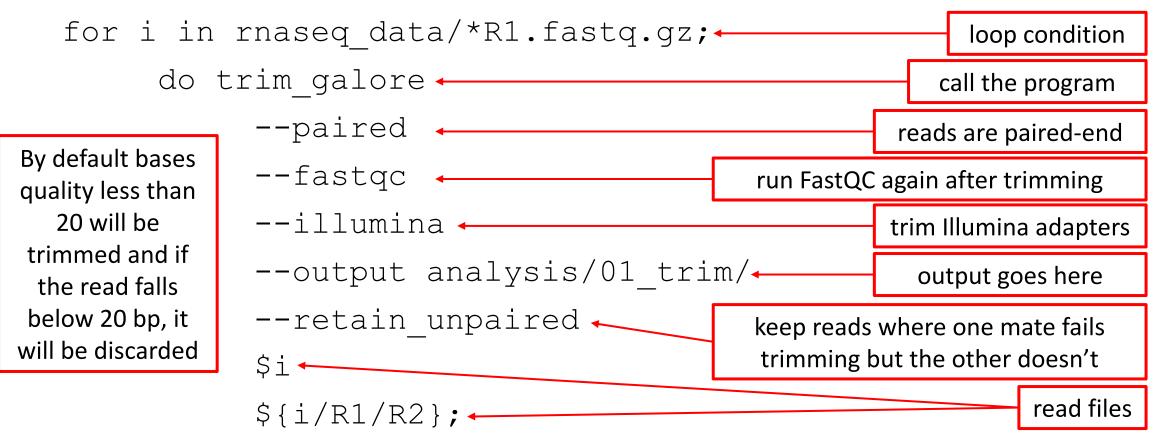










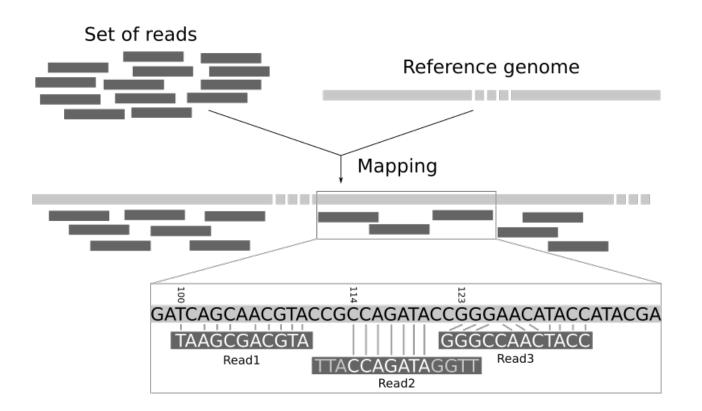


for i in rnaseq_data/*R1.fastq.gz; do trim_galore
--paired --fastqc --illumina --output analysis/01_trim/
--retain unpaired \$i \${i/R1/R2}; done

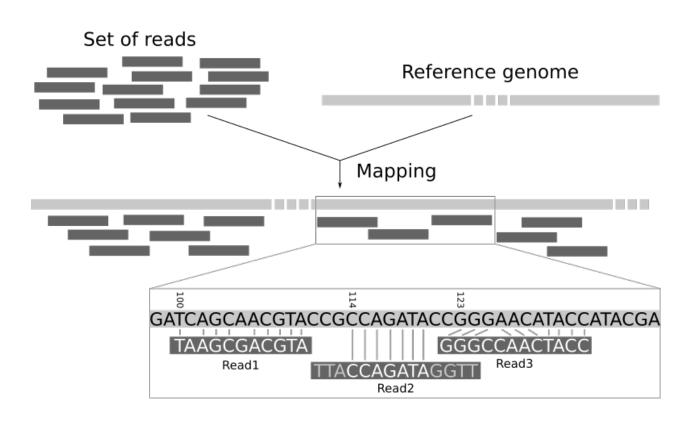
Align

How does aligning work?

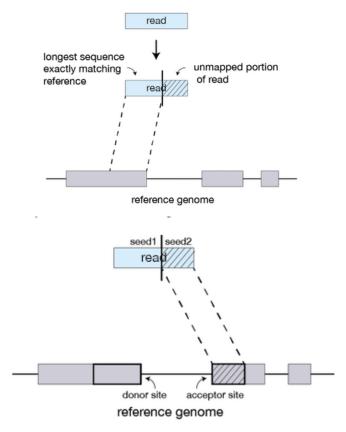
How does aligning work?



How does aligning work?



STAR (Spliced Transcripts Alignment to a Reference)



- 1. Make a folder inside the analysis folder to put the aligned reads in, analysis/02_align
- 2. Change to the trimmed reads folder analysis/01_trim

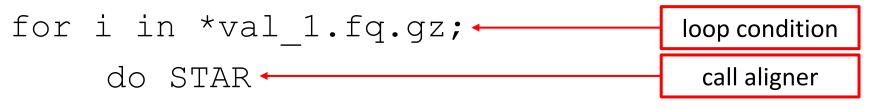
- - /hg38/chr21/STAR_index --readFilesIn \$i \${i/R1_val_1/R2_val_2}

--readFilesCommand zcat

--outFileNamePrefix ../02_align/\${i/R1*/}

--outSAMtype BAM SortedByCoordinate;

for i in *val_1.fq.gz; Ioop condition
do STAR



--genomeDir /mnt/data/gdata/human \

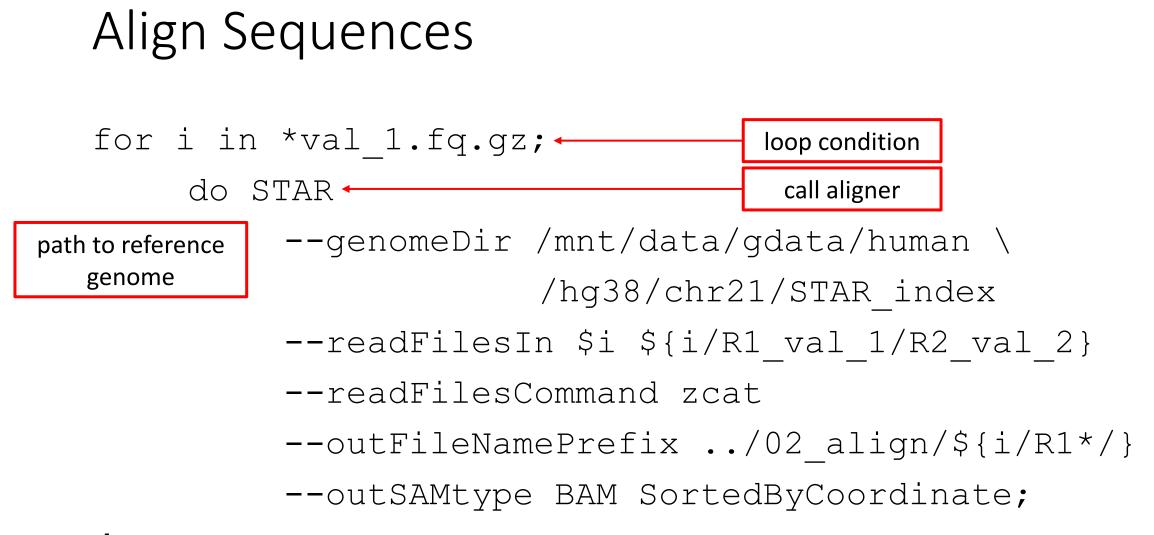
/hg38/chr21/STAR_index

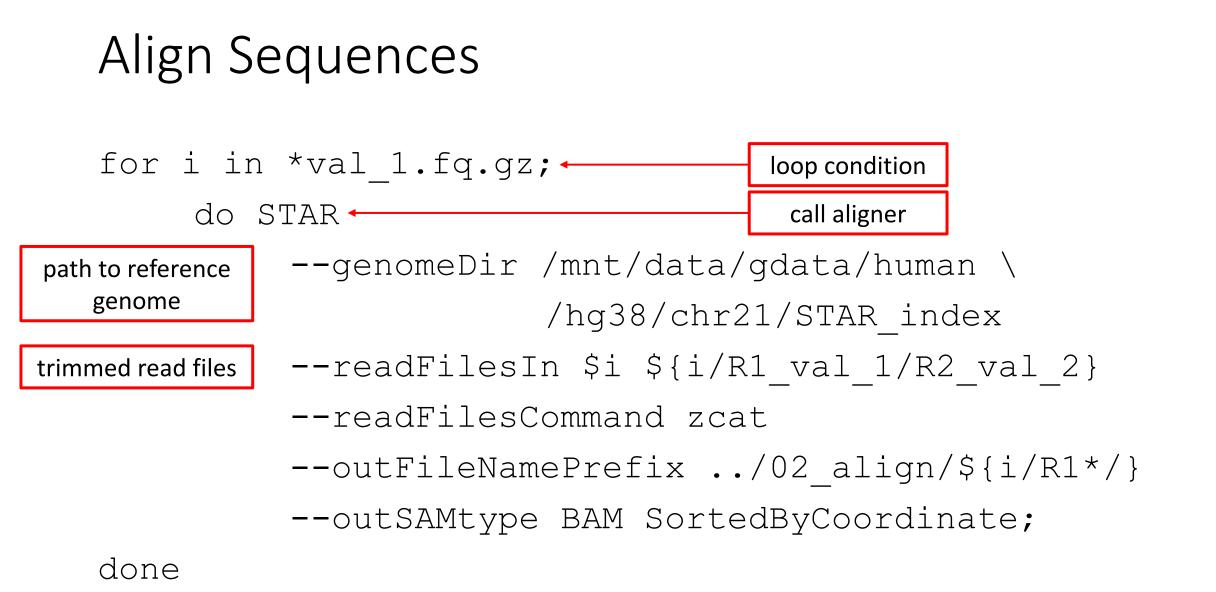
--readFilesIn \$i \${i/R1_val_1/R2_val_2}

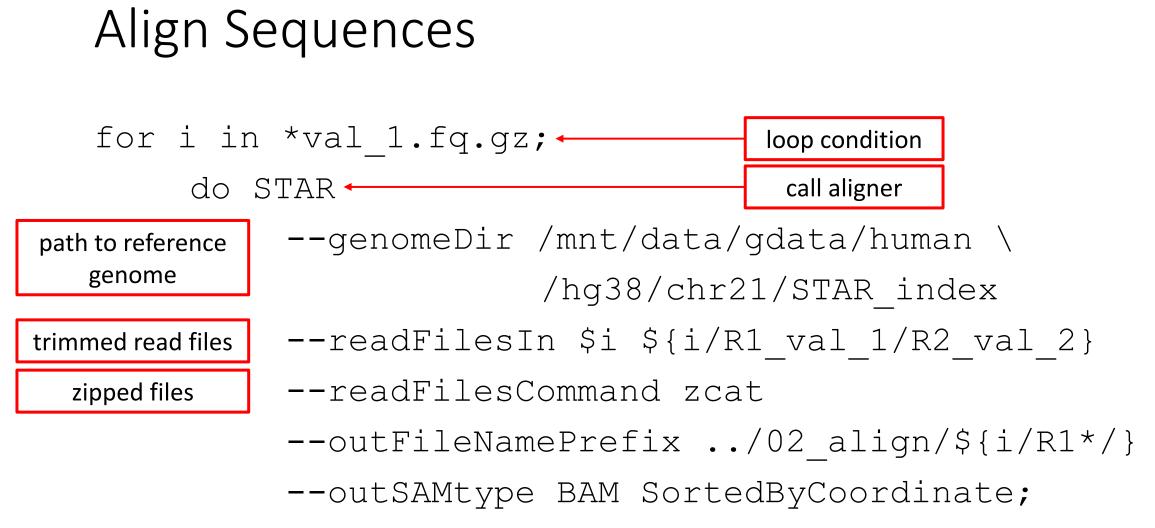
--readFilesCommand zcat

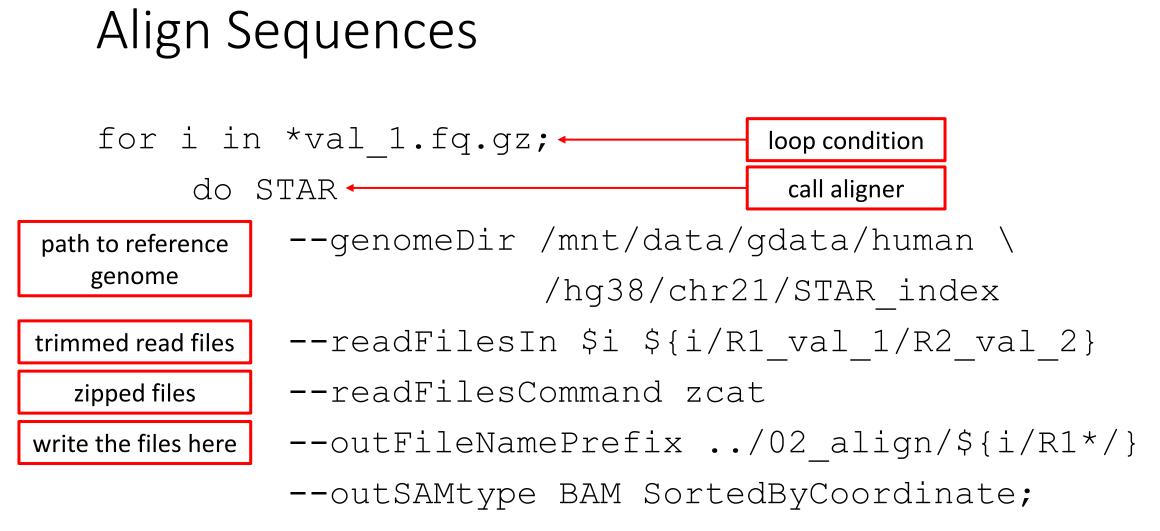
--outFileNamePrefix ../02 align/\${i/R1*/}

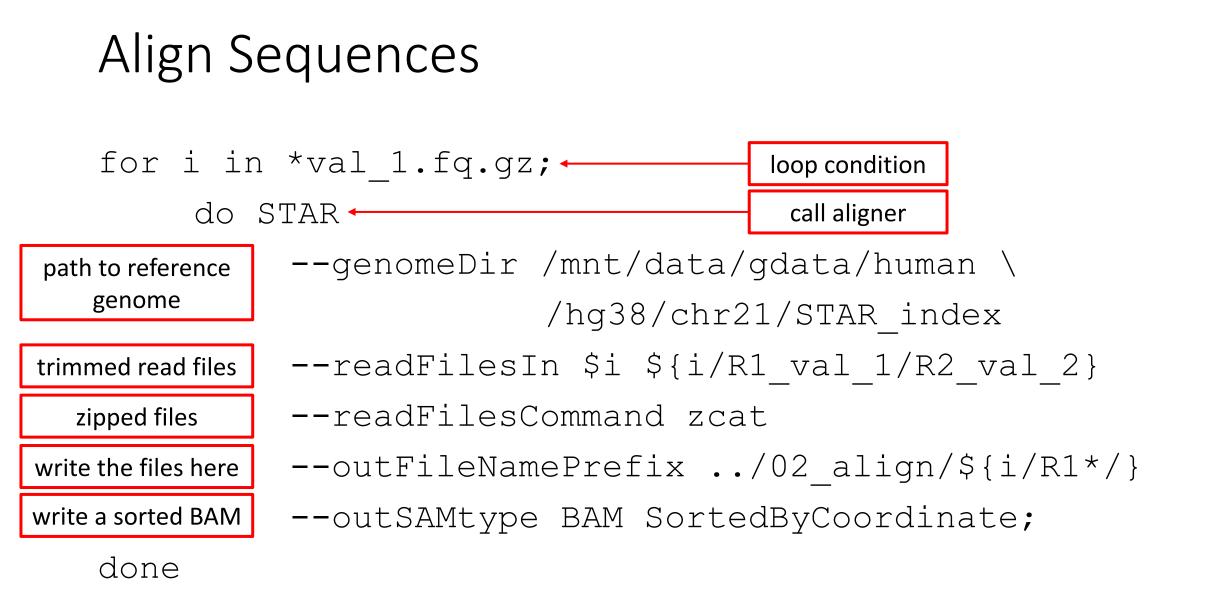
--outSAMtype BAM SortedByCoordinate;











Align Sequences

for i in *val_1.fq.gz; do STAR --genomeDir /mnt/data/gdata/human/hg38/chr21/STAR_index -readFilesIn \$i \${i/R1_val_1/R2_val_2} -readFilesCommand zcat --outFileNamePrefix ../02_align/\${i/R1*/} --outSAMtype BAM SortedByCoordinate; done

What do you mean by count features?

- We're going to count genes, but you could also count:
 - transcripts
 - non-coding RNA
- Need an annotation file for whatever feature you want to count

<u>Col 1</u>	Col 2	Col 3	Col 4	<u>Col 5</u>	Col 6	Col 7	Col 8	<u>Col 9</u>
chr21	HAVANA	transcript	10862622	10863067	. •	+		gene id "ENSG00000169
chr21	HAVANA	exon	10862622	10862667		+		gene id "ENSG00000169
chr21	HAVANA	CDS	10862622	10862667		+	0	gene id "ENSG00000169
chr21	HAVANA	start codon	10862622	10862624		+	0	gene id "ENSG00000169
chr21	HAVANA	exon -	10862751	10863067		+		gene id "ENSG00000169
chr21	HAVANA	CDS	10862751	10863064		+	2	gene id "ENSG00000169
chr21	HAVANA	stop codon	10863065	10863067		+	0	gene id "ENSG00000169
chr21	HAVANA	UTR	10863065	10863067		+		gene_id "ENSG00000169

• Going to use a gene transfer format (GTF) file for annotations

- 1. Make a folder inside the analysis folder to put the aligned reads in, .../03_count
- 2. Change to the trimmed reads folder .../02_align/

for i in *.bam;

do featureCounts

-a /mnt/data/gdata/human/hg38/chr21/ \
 homo_sapiens_hg38_chr21.gtf
-o ../03_count/\${i/ \
 Aligned.sortedByCoord.out.bam/ \
 counts.txt}
-R BAM

\$i;

for i in *.bam; -

loop condition

do featureCounts

-a /mnt/data/gdata/human/hg38/chr21/ \

homo_sapiens_hg38_chr21.gtf

-o $../03_count/$ \${i/ \

Aligned.sortedByCoord.out.bam/ $\$

counts.txt}

-R BAM

\$i;

for i in *.bam; loop condition do featureCounts + call program -a /mnt/data/gdata/human/hg38/chr21/ \ homo sapiens hg38 chr21.gtf -o ../03 count/\${i/ \ Aligned.sortedByCoord.out.bam/ \ counts.txt} -R BAM

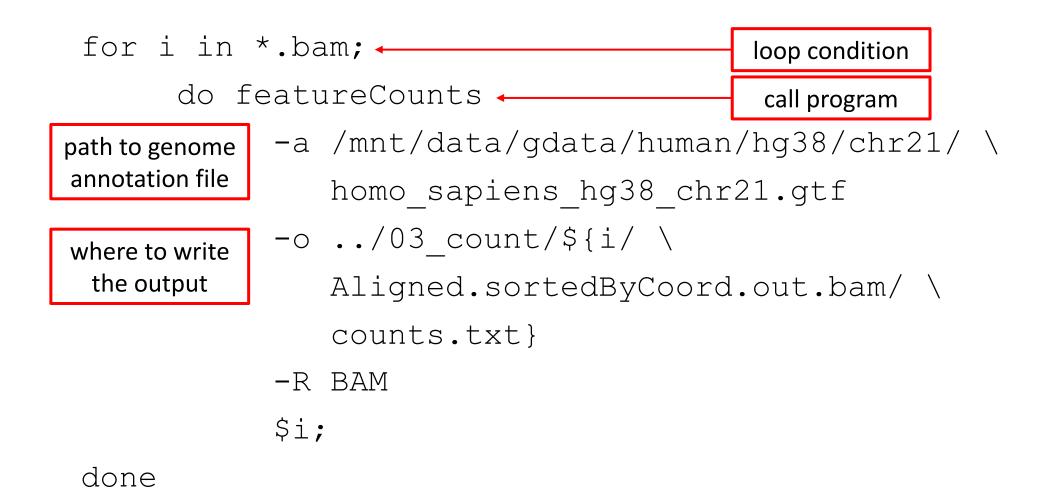
for i in *.bam; -

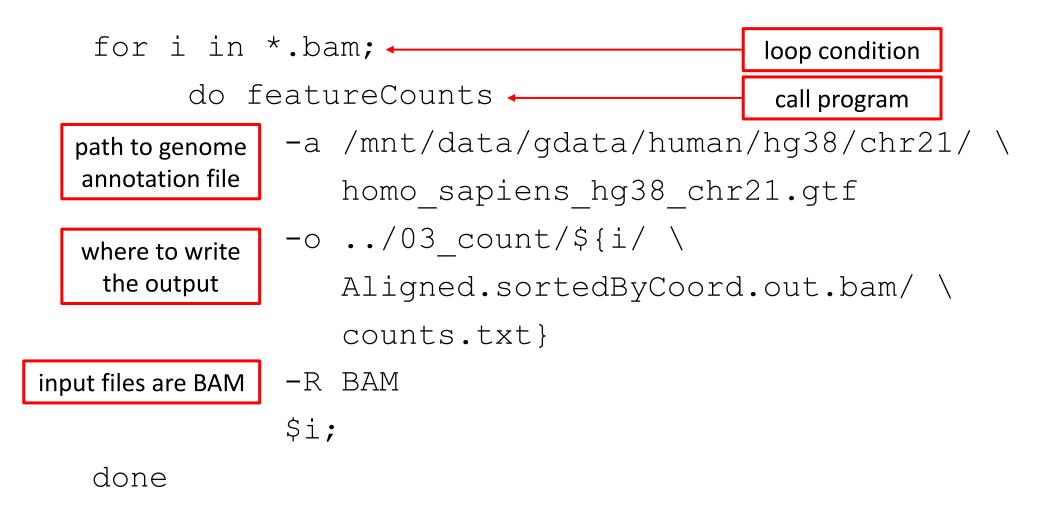
do featureCounts 🔶

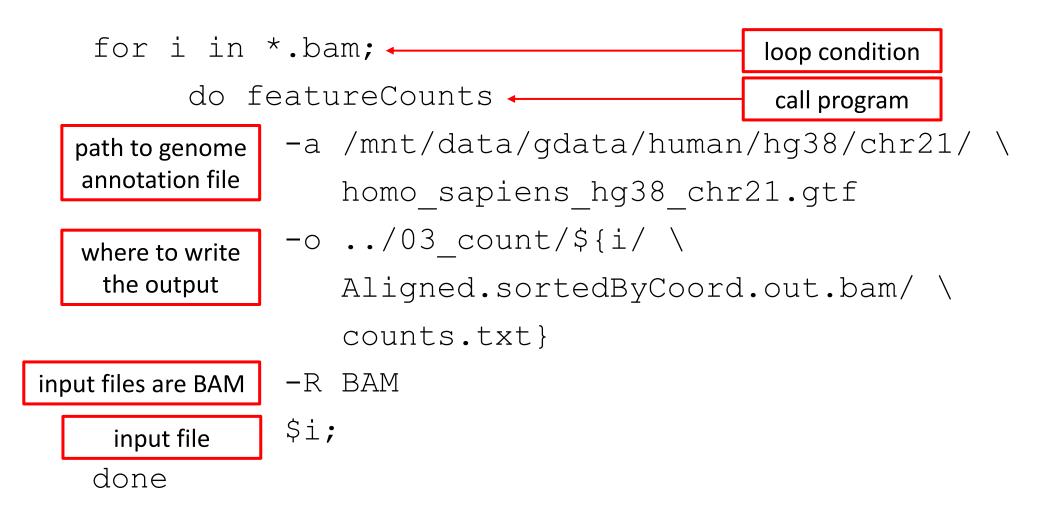
loop condition call program

path to genome annotation file

- -a /mnt/data/gdata/human/hg38/chr21/ \
 homo_sapiens_hg38_chr21.gtf
 -o ../03_count/\${i/ \
 Aligned.sortedByCoord.out.bam/ \
 counts.txt}
 -R BAM
- \$i;







for i in *.bam; do featureCounts -a
/mnt/data/gdata/human/hg38/chr21/homo_sapiens_hg
38_chr21.gtf -o

../03_count/\${i/Aligned.sortedByCoord.out.bam/co
unts.txt} -R BAM \$i; done

General Steps

- 1. Check quality
- 2. Trim
- 3. Align
- 4. Count features
- 5. Statistics