

Register= <https://imjo.in/qw9eX3>

SNP calling , SAM-representation of variants, OMIM Lookup, Alignment Pileups and VCF format, Annotation and filtering of VCF files using ANNOVAR- Examples, Variant calling on sample data, Multi-sample Variant calling, Variant Normalization and Variant Effect prediction using VEP (Variant Effect Predictor)/ SnpEff.

Contact NGS Instructor: +91 93 511 755 63
<https://www.linkedin.com/in/kspsharma/>

sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible: ~/dc_workshop

(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/dc_workshop\$ samtools tview results/bam/SRR2584866.aligned.sorted.bam data/ref_genome/ecoli.fa

(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/dc_workshop\$ ls -lh /*

total 345M

-rw-rw-r-- 1 sharmaimr sharmaimr 4.5M Sep 1 11:44 ecoli.fa
-rw-rw-r-- 1 sharmaimr sharmaimr 12 Sep 1 11:54 ecoli.fa.amb
-rw-rw-r-- 1 sharmaimr sharmaimr 86 Sep 1 11:54 ecoli.fa.ann
-rw-rw-r-- 1 sharmaimr sharmaimr 4.5M Sep 1 11:54 ecoli.fa.bwt
-rw-rw-r-- 1 sharmaimr sharmaimr 28 Sep 1 12:12 ecoli.fa.fai
-rw-rw-r-- 1 sharmaimr sharmaimr 1.2M Sep 1 11:54 ecoli.fa.pac
-rw-rw-r-- 1 sharmaimr sharmaimr 2.3M Sep 1 11:54 ecoli.fa.sa
-rw-rw-r-- 1 sharmaimr sharmaimr 1.4M Sep 1 11:42 GCA_000017985.1_ASM1798v1_genomic.fna.gz
-rw-rw-r-- 1 sharmaimr sharmaimr 332M Sep 1 11:48 sub.tar

data/trimmed_fastq_small:

total 332M

-rw-rw-r-- 1 sharmaimr sharmaimr 58M Jul 7 2018 SRR2584863_1.trim.sub.fastq
-rw-rw-r-- 1 sharmaimr sharmaimr 53M Jul 7 2018 SRR2584863_2.trim.sub.fastq
-rw-rw-r-- 1 sharmaimr sharmaimr 55M Jul 7 2018 SRR2584866_1.trim.sub.fastq
-rw-rw-r-- 1 sharmaimr sharmaimr 57M Jul 7 2018 SRR2584866_2.trim.sub.fastq
-rw-rw-r-- 1 sharmaimr sharmaimr 58M Jul 7 2018 SRR2589044_1.trim.sub.fastq
-rw-rw-r-- 1 sharmaimr sharmaimr 53M Jul 7 2018 SRR2589044_2.trim.sub.fastq

results/bam:

total 73M

-rw-rw-r-- 1 sharmaimr sharmaimr 42M Sep 1 12:08 SRR2584866.aligned.bam
-rw-rw-r-- 1 sharmaimr sharmaimr 31M Sep 1 12:08 SRR2584866.aligned.sorted.bam
-rw-rw-r-- 1 sharmaimr sharmaimr 13K Sep 1 12:22 SRR2584866.aligned.sorted.bam.bai

results/bcf:

total 75M

-rw-rw-r-- 1 sharmaimr sharmaimr 74M Sep 1 12:14 SRR2584866_raw.bcf
-rw-rw-r-- 1 sharmaimr sharmaimr 99K Sep 1 12:17 SRR2584866_variants.vcf

results/sam:

total 125M

-rw-rw-r-- 1 sharmaimr sharmaimr 125M Sep 1 12:06 SRR2584866.aligned.sam

results/vcf:

total 100K

-rw-rw-r-- 1 sharmaimr sharmaimr 98K Sep 1 12:18 SRR2584866_final_variants.vcf

(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/dc_workshop\$ █

[File](#) [Genomes](#) [View](#) [Tracks](#) [Regions](#) [Tools](#) [GenomeSpace](#) [Help](#)

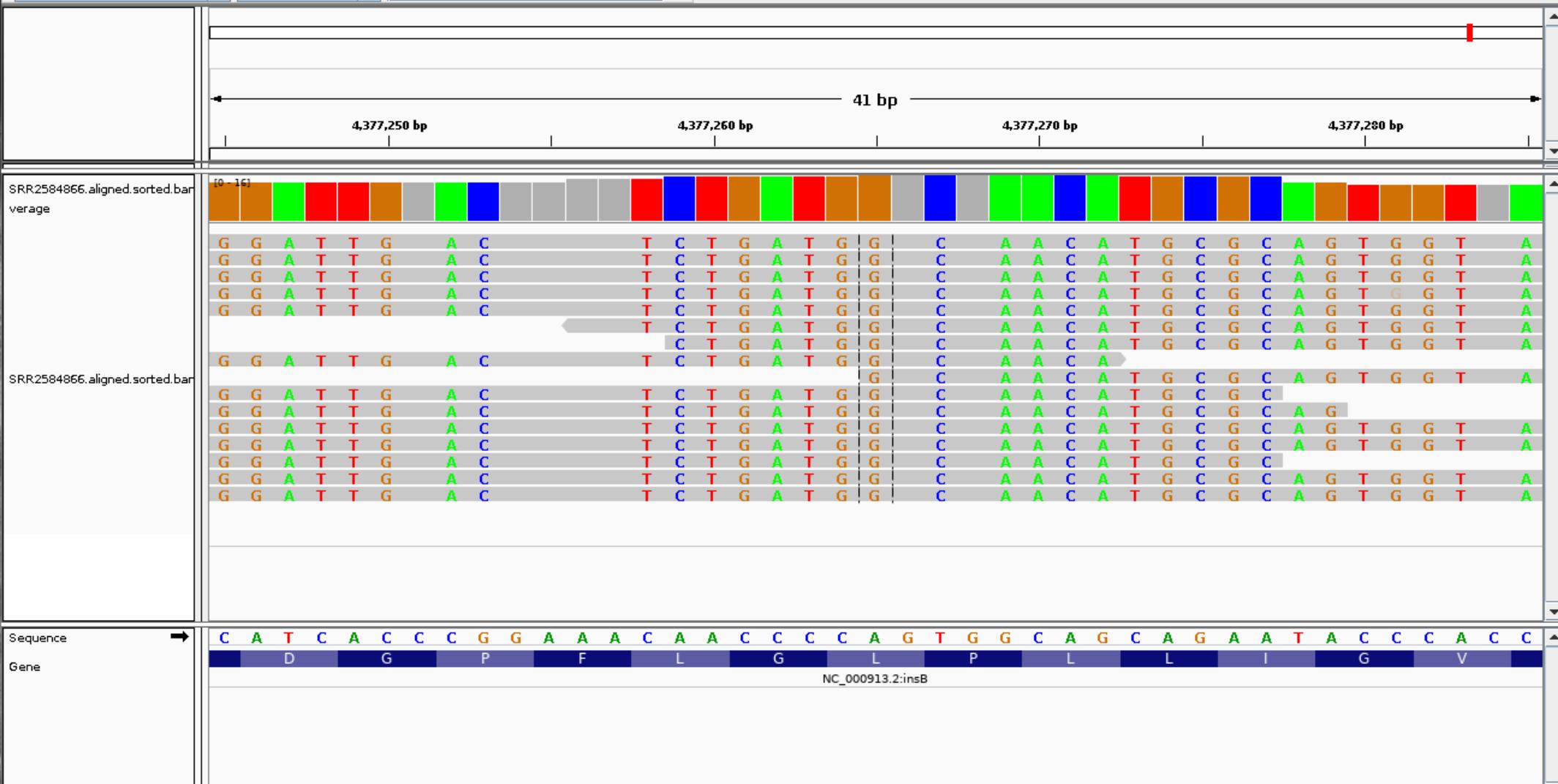
E. coli K-12 MG1655 ...

U00096.2

U00096.2:4,377,245-4,377,285



[-] [+]



```
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ls  
snpecoli.pdf  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ mkdir -p data results  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ cd data  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data$ ls  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data$ mkdir ref_genome  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data$ ls  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ efet  
ch -db nucleotide -id NC_000913.3 --fo  
rmat fasta | head  
>NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome  
AGCTTTCACTCTGACTGCAACGGGCAATATGTCTCTGTGATTAAAAAAAGAGTGTCTGATAGCAGC  
TTCTGAACCTGGTACCTGCCGTGAGTAATTAAAATTATTGACTTAGGTCACTAAATCTTAACCAA  
TATAGGCATAGCGCACAGACAGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACC  
ATTACCACCCATCACCATACCCATTACCGTAACGGTGCGGGCTGACGGTACAGGAAACACAGAAAAAAG  
CCCGCACCTGACAGTGGCGTTTTTGACCAAAGGTAACGAGGTAACAACCATGCGAGTGTGAA  
GTTGGCGGTACATCAGTGGCAAATGAGAACGTTTCTGCGTGTGCGATATTCTGAAAGCAATGCC  
AGGCAGGGCAGGTGGCCACCGTCTCTGCCCCGCCAAATCACAACCCATTGGTGGCGATGATTG  
AAAAAACATTAGCGGCCAGGATGCTTACCAAATATCAGCGATGCCAACGTATTTGCGGAACCTT  
GACGGGACTCGCCGCCAGCCGGTCTCCGCTGGCGCAATTGAAAACTTCGTCGATCAGGAATT  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ efet  
ch -db nucleotide -id NC_000913.3 --fo  
rmat fasta > ecoli.fa  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ seqkit stat ecoli.fa  
file format type num_seqs sum_len min_len avg_len max_len  
ecoli.fa FASTA DNA 1 4,641,652 4,641,652 4,641,652 4,641,652  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ bwa index ecoli.fa  
[bwa_index] Pack FASTA... 0.10 sec  
[bwa_index] Construct BWT for the packed sequence...  
[bwa_index] 7.20 seconds elapse.  
[bwa_index] Update BWT... 0.08 sec  
[bwa_index] Pack forward-only FASTA... 0.06 sec  
[bwa_index] Construct SA from BWT and Occ... 2.84 sec  
[main] Version: 0.7.17-r1188  
[main] CMD: bwa index ecoli.fa  
[main] Real time: 10.883 sec; CPU: 10.281 sec  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ █
```

```
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ curl -L -o sub.tar.gz https://ndownloader.figshare.com/files/14418248
% Total    % Received % Xferd  Average Speed   Time   Time     Time  Current
                                         Dload  Upload   Total   Spent   Left  Speed
0       0      0      0      0      0      0  --::--:--  0:00:01  --::--:--      0
100  109M  100  109M      0      0  1662k      0  0:01:07  0:01:07  --::--:-- 2321k
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ ls
ecoli.fa ecoli.fa.amb ecoli.fa.ann ecoli.fa.bwt ecoli.fa.pac ecoli.fa.sa sub.tar.gz
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ tar -xzf sub.tar.gz
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ ls
ecoli.fa ecoli.fa.amb ecoli.fa.ann ecoli.fa.bwt ecoli.fa.pac ecoli.fa.sa sub sub.tar.gz
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ mv sub/ ./trimmed_fastq_small
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ ls
ecoli.fa ecoli.fa.amb ecoli.fa.ann ecoli.fa.bwt ecoli.fa.pac ecoli.fa.sa sub.tar.gz trimmed_fastq_small
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ cd trimmed_fastq_small/
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome/trimmed_fastq_small$ ls -lh *
-rw-rw-r-- 1 sharmaimr sharmaimr 58M Jul  7 2018 SRR2584863_1.trim.sub.fastq
-rw-rw-r-- 1 sharmaimr sharmaimr 53M Jul  7 2018 SRR2584863_2.trim.sub.fastq
-rw-rw-r-- 1 sharmaimr sharmaimr 55M Jul  7 2018 SRR2584866_1.trim.sub.fastq
-rw-rw-r-- 1 sharmaimr sharmaimr 57M Jul  7 2018 SRR2584866_2.trim.sub.fastq
-rw-rw-r-- 1 sharmaimr sharmaimr 58M Jul  7 2018 SRR2589044_1.trim.sub.fastq
-rw-rw-r-- 1 sharmaimr sharmaimr 53M Jul  7 2018 SRR2589044_2.trim.sub.fastq
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome/trimmed_fastq_small$ cd ..
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ ls
ecoli.fa ecoli.fa.amb ecoli.fa.ann ecoli.fa.bwt ecoli.fa.pac ecoli.fa.sa sub.tar.gz trimmed_fastq_small
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ cd ..
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data$ ls
ref_genome
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data$ cd ..
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ls
data results snpecoli.pdf
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ mkdir -p results/sam results/bam results/bcf results/vcf
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ls ./results/
bam bcf sam vcf
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$
```

sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible: ~/ngs/snp/snp01sep2020/data/ref_genome

```
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ls
data results snpecoli.pdf
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ cd data
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data$ ls
ref_genome
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data$ cd ref_genome/
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ ls
ecoli.fa ecoli.fa.amb ecoli.fa.ann ecoli.fa.bwt ecoli.fa.pac ecoli.fa.sa sub.tar.gz trimmed_fastq_small
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ head -n2 ecoli.fa
>NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome
AGCTTTTCAATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGC
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ mkdir backup
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ mv ecoli.* ./backup/
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ ls
backup ecoli.fa sub.tar.gz trimmed_fastq_small
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ nano ecoli.fa
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ head ecoli.fa
>NC_000913.2
AGCTTTTCAATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGC
TTCTGAACCTGGTACCTGCCGTGAGTAATTAATTTATTGACTTAGGTACTAAATACTTAACCAA
TATAGGCATAGCGCACAGACAGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACC
ATTACCACCAACATCACCATTACCAACAGGTAAACGGTGCGGGCTGACGGTACAGGAAACACAGAAAAAAG
CCCGCACCTGACAGTGGGGCTTTTTTCGACCAAGGTAAACGAGGTAAACAACCATGCGAGTGTGAA
GTTGGCGGTACATCAGTGGCAAATGCAGAACGTTTCTGCGTGTGCGATATTCTGAAAGCAATGCC
AGGCAGGGGCAGGTGCCACCGTCTCTGCCCCGCCAAATCACCAACCACCTGGTGGCGATGATTG
AAAAAACCATTAGCGGCCAGGATGCTTACCCAATATCAGCGATGCCAACGTATTTGCCAACACTTT
GACGGGACTCGCCGCCAGCGGGGTTCCCGTGGCGCAATTGAAAACTTCGTCATCAGGAATTG
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ bwa index ecoli.fa
[bwa_index] Pack FASTA... 0.14 sec
[bwa_index] Construct BWT for the packed sequence...
[bwa_index] 8.17 seconds elapse.
[bwa_index] Update BWT... 0.11 sec
[bwa_index] Pack forward-only FASTA... 0.10 sec
[bwa_index] Construct SA from BWT and Occ... 3.37 sec
[main] Version: 0.7.17-r1188
[main] CMD: bwa index ecoli.fa
[main] Real time: 12.270 sec; CPU: 11.911 sec
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ █
```

```

backup ecoli.fa ecoli.fa.amb ecoli.fa.ann ecoli.fa.bwt ecoli.fa.pac ecoli.fa.sa sub.tar.gz trimmed_fastq_small
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ bwa mem ecoli.fa trimmed_fastq_small/SRR25
84866_1.trim.sub.fastq trimmed_fastq_small/SRR2584866_2.trim.sub.fastq | head -n5
[M::bwa_idx_load_from_disk] read 0 ALT contigs
[M::process] read 77446 sequences (10000033 bp)...
[M::process] read 77296 sequences (10000182 bp)...
[M::mem_pestat] # candidate unique pairs for (FF, FR, RF, RR): (46, 35173, 21, 56)
[M::mem_pestat] analyzing insert size distribution for orientation FF...
[M::mem_pestat] (25, 50, 75) percentile: (384, 692, 1774)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 4554)
[M::mem_pestat] mean and std.dev: (827.05, 750.19)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 5944)
[M::mem_pestat] analyzing insert size distribution for orientation FR...
[M::mem_pestat] (25, 50, 75) percentile: (220, 359, 576)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 1288)
[M::mem_pestat] mean and std.dev: (411.81, 227.75)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 1644)
[M::mem_pestat] analyzing insert size distribution for orientation RF...
[M::mem_pestat] (25, 50, 75) percentile: (1008, 2173, 2637)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 5895)
[M::mem_pestat] mean and std.dev: (1811.85, 1226.27)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 7524)
[M::mem_pestat] analyzing insert size distribution for orientation RR...
[M::mem_pestat] (25, 50, 75) percentile: (324, 563, 942)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 2178)
[M::mem_pestat] mean and std.dev: (592.36, 433.09)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 2796)
[M::mem_pestat] skip orientation FF
[M::mem_pestat] skip orientation RF
[M::mem_pestat] skip orientation RR
[M::mem_process_seqs] Processed 77446 reads in 21.014 CPU sec, 20.800 real sec
@SQ SN:NC_000913.2 LN:4641652
@PG ID:bwa PN:bwa VN:0.7.17-r1188 CL:bwa mem ecoli.fa trimmed_fastq_small/SRR2584866_1.trim.sub.fastq trimmed_fastq_small/SRR2584866_2.tr
im.sub.fastq
SRR2584866.7 83 NC_000913.2 3655777 60 26M = 3655152 -651 GCTTCCCATGCCAATTAAATACATGTG IJJJJJJJJJJJJHHHHHFFFFF
CC@ NM:i:0 MD:Z:26 MC:Z:150M AS:i:26 XS:i:0
SRR2584866.7 163 NC_000913.2 3655152 60 150M = 3655777 651 TCCTCAATCTTAAGCGGATCAATGAGCTGGTACGCCATCAGCATATTGATTATCT
GGTGTGAATTTCAGGCTTACGGTGGACTGGCTACGCTGCCACACAGATTAGCTAATTGAAACGCCCTTCACCCCTGCCATACCTTTAATAATC C@FFFFFGHHGGGGJJJJJJJJJJJIHIIJJJJJJJJJJJJ
IIIJJJJJJJJJJHHHHHHFFFFFDDDDDDDDDDDDDBBDDDDDDDEDEDDEDDDDDDDBBDDCDDDDDBBBDDDDDDDEDED NM:i:0 MD:Z:150 MC:Z:26M A
XS:i:150 XS:i:0
SRR2584866.8 99 NC_000913.2 2967216 60 26M = 2967713 647 GCTGATATTCTGCAGCAGTACCGGC @@@DBDDDHFFHIFFGBH<FHC
GDG NM:i:0 MD:Z:26 MC:Z:150M AS:i:26 XS:i:0
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ 
```

```
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ ls
backup ecoli.fa ecoli.fa.amb ecoli.fa.ann ecoli.fa.bwt ecoli.fa.pac ecoli.fa.sa sub.tar.gz trimmed_fastq_small
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ ls ../../results/sam/
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$ bwa mem ecoli.fa trimmed_fastq_small/SRR25
84866_1.trim.sub.fastq trimmed_fastq_small/SRR2584866_2.trim.sub.fastq > ../../results/sam/SRR2584866.aligned.sam
[M::bwa_idx_load_from_disk] read 0 ALT contigs
[M::process] read 77446 sequences (10000033 bp)...
[M::process] read 77296 sequences (10000182 bp)...
[M::mem_pestat] # candidate unique pairs for (FF, FR, RF, RR): (46, 35173, 21, 56)
[M::mem_pestat] analyzing insert size distribution for orientation FF...
[M::mem_pestat] (25, 50, 75) percentile: (384, 692, 1774)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 4554)
[M::mem_pestat] mean and std.dev: (827.05, 750.19)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 5944)
[M::mem_pestat] analyzing insert size distribution for orientation FR...
[M::mem_pestat] (25, 50, 75) percentile: (220, 359, 576)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 1288)
[M::mem_pestat] mean and std.dev: (411.81, 227.75)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 1644)
[M::mem_pestat] analyzing insert size distribution for orientation RF...
[M::mem_pestat] (25, 50, 75) percentile: (1008, 2173, 2637)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 5895)
[M::mem_pestat] mean and std.dev: (1811.85, 1226.27)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 7524)
[M::mem_pestat] analyzing insert size distribution for orientation RR...
[M::mem_pestat] (25, 50, 75) percentile: (324, 563, 942)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 2178)
[M::mem_pestat] mean and std.dev: (592.36, 433.09)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 2796)
[M::mem_pestat] skip orientation FF
[M::mem_pestat] skip orientation RF
[M::mem_pestat] skip orientation RR
[M::mem_process_seqs] Processed 77446 reads in 21.042 CPU sec, 20.796 real sec
[M::process] read 75458 sequences (10000048 bp)...
[M::mem_pestat] # candidate unique pairs for (FF, FR, RF, RR): (64, 35075, 11, 59)
[M::mem_pestat] analyzing insert size distribution for orientation FF...
[M::mem_pestat] (25, 50, 75) percentile: (172, 460, 1166)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 3154)
[M::mem_pestat] mean and std.dev: (654.62, 712.78)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 4148)
[M::mem_pestat] analyzing insert size distribution for orientation FR...
[M::mem_pestat] (25, 50, 75) percentile: (223, 367, 581)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 1297)
```

```
[M::mem_pestat] (25, 50, 75) percentile: (223, 367, 581)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 1297)
[M::mem_pestat] mean and std.dev: (416.81, 229.17)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 1655)
[M::mem_pestat] analyzing insert size distribution for orientation RF...
[M::mem_pestat] (25, 50, 75) percentile: (2317, 2349, 2422)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (2107, 2632)
[M::mem_pestat] mean and std.dev: (2319.12, 75.27)
[M::mem_pestat] low and high boundaries for proper pairs: (2002, 2737)
[M::mem_pestat] analyzing insert size distribution for orientation RR...
[M::mem_pestat] (25, 50, 75) percentile: (244, 472, 1366)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 3610)
[M::mem_pestat] mean and std.dev: (734.04, 824.81)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 4732)
[M::mem_pestat] skip orientation FF
[M::mem_pestat] skip orientation RF
[M::mem_pestat] skip orientation RR
[M::mem_process_seqs] Processed 77296 reads in 21.216 CPU sec, 21.054 real sec
[M::process] read 74054 sequences (10000155 bp)...
[M::mem_pestat] # candidate unique pairs for (FF, FR, RF, RR): (51, 34174, 14, 74)
[M::mem_pestat] analyzing insert size distribution for orientation FF...
[M::mem_pestat] (25, 50, 75) percentile: (288, 501, 1088)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 2688)
[M::mem_pestat] mean and std.dev: (653.18, 611.90)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 3488)
[M::mem_pestat] analyzing insert size distribution for orientation FR...
[M::mem_pestat] (25, 50, 75) percentile: (223, 360, 575)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 1279)
[M::mem_pestat] mean and std.dev: (413.55, 226.84)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 1631)
[M::mem_pestat] analyzing insert size distribution for orientation RF...
[M::mem_pestat] (25, 50, 75) percentile: (643, 2230, 2621)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 6577)
[M::mem_pestat] mean and std.dev: (1685.64, 977.13)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 8555)
[M::mem_pestat] analyzing insert size distribution for orientation RR...
[M::mem_pestat] (25, 50, 75) percentile: (365, 611, 1831)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 4763)
[M::mem_pestat] mean and std.dev: (1120.91, 1214.32)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 6229)
[M::mem_pestat] skip orientation FF
[M::mem_pestat] skip orientation RF
[M::mem_pestat] skip orientation RR
```

sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible: ~/ngs/snp/snp01sep2020/data/ref_genome

```
[M::mem_pestat] analyzing insert size distribution for orientation FR...
[M::mem_pestat] (25, 50, 75) percentile: (219, 352, 565)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 1257)
[M::mem_pestat] mean and std.dev: (406.83, 223.23)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 1603)
[M::mem_pestat] analyzing insert size distribution for orientation RF...
[M::mem_pestat] (25, 50, 75) percentile: (143, 2145, 2406)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 6932)
[M::mem_pestat] mean and std.dev: (1205.16, 1101.74)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 9195)
[M::mem_pestat] analyzing insert size distribution for orientation RR...
[M::mem_pestat] (25, 50, 75) percentile: (214, 400, 918)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 2326)
[M::mem_pestat] mean and std.dev: (519.02, 526.86)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 3030)
[M::mem_pestat] skip orientation FF
[M::mem_pestat] skip orientation RF
[M::mem_pestat] skip orientation RR
[M::mem_process_seqs] Processed 74054 reads in 21.296 CPU sec, 21.376 real sec
[M::mem_pestat] # candidate unique pairs for (FF, FR, RF, RR): (36, 20737, 9, 30)
[M::mem_pestat] analyzing insert size distribution for orientation FF...
[M::mem_pestat] (25, 50, 75) percentile: (273, 846, 1382)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 3600)
[M::mem_pestat] mean and std.dev: (801.59, 685.63)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 4709)
[M::mem_pestat] analyzing insert size distribution for orientation FR...
[M::mem_pestat] (25, 50, 75) percentile: (218, 349, 559)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 1241)
[M::mem_pestat] mean and std.dev: (402.95, 221.42)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 1582)
[M::mem_pestat] skip orientation RF as there are not enough pairs
[M::mem_pestat] analyzing insert size distribution for orientation RR...
[M::mem_pestat] (25, 50, 75) percentile: (240, 557, 708)
[M::mem_pestat] low and high boundaries for computing mean and std.dev: (1, 1644)
[M::mem_pestat] mean and std.dev: (505.89, 314.27)
[M::mem_pestat] low and high boundaries for proper pairs: (1, 2112)
[M::mem_pestat] skip orientation FF
[M::mem_pestat] skip orientation RR
[M::mem_process_seqs] Processed 45746 reads in 13.166 CPU sec, 12.922 real sec
[main] Version: 0.7.17-r1188
[main] CMD: bwa mem ecoli.fa trimmed_fastq_small/SRR2584866_1.trim.sub.fastq trimmed_fastq_small/SRR2584866_2.trim.sub.fastq
[main] Real time: 98.254 sec; CPU: 98.698 sec
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/data/ref_genome$
```

```
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ samtools view -S -b results/sam/SRR2584866.aligned.sam > results/bam/SRR2584866.aligned.bam
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ samtools sort -o results/bam/SRR2584866.aligned.sorted.bam results/bam/SRR2584866.aligned.bam
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ samtools index results/bam/SRR2584866.aligned.sorted.bam
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ls -lh results/sam
total 125M
-rw-rw-r-- 1 sharmaimr sharmaimr 125M Sep  1 15:15 SRR2584866.aligned.sam
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ls -lh results/bam/
total 75M
-rw-rw-r-- 1 sharmaimr sharmaimr 43M Sep  1 15:20 SRR2584866.aligned.bam
-rw-rw-r-- 1 sharmaimr sharmaimr 32M Sep  1 15:21 SRR2584866.aligned.sorted.bam
-rw-rw-r-- 1 sharmaimr sharmaimr 13K Sep  1 15:21 SRR2584866.aligned.sorted.bam.bai
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ samtools flagstat results/bam/bcf/
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ samtools flagstat results/bam/SRR2584866.aligned.bam
351753 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
1753 + 0 supplementary
0 + 0 duplicates
336477 + 0 mapped (95.66% : N/A)
350000 + 0 paired in sequencing
175000 + 0 read1
175000 + 0 read2
328580 + 0 properly paired (93.88% : N/A)
333166 + 0 with itself and mate mapped
1558 + 0 singletons (0.45% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ<=5)
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ samtools flagstat results/bam/SRR2584866.aligned.sorted.bam
351753 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
1753 + 0 supplementary
0 + 0 duplicates
336477 + 0 mapped (95.66% : N/A)
350000 + 0 paired in sequencing
175000 + 0 read1
175000 + 0 read2
328580 + 0 properly paired (93.88% : N/A)
333166 + 0 with itself and mate mapped
1558 + 0 singletons (0.45% : N/A)
0 + 0 with mate mapped to a different chr
```

```
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ls
data results snpecoli.pdf
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ bcftools mpileup -O b -o results/bcf/SRR2584866_raw.bcf -f
data/ref_genome/ecoli.fa results/bam/SRR2584866.aligned.sorted.bam
[mpileup] 1 samples in 1 input files
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ bcftools call --ploidy 1 -m -v -o results/bcf/SRR2584866_v
ariants.vcf results/bcf/SRR2584866_raw.bcf
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ls
data results snpecoli.pdf
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ vcfutils.pl varFilter results/bcf/SRR2584866_variants.vcf
| head -n35 results/vcf/SRR2584866_final_variants.vcf
head: cannot open 'results/vcf/SRR2584866_final_variants.vcf' for reading: No such file or directory
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ vcfutils.pl varFilter results/bcf/SRR2584866_variants.vcf
> results/vcf/SRR2584866_final_variants.vcf
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ head -n35 results/vcf/SRR2584866_final_variants.vcf
##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters passed">
##bcftoolsVersion=1.9+htslib-1.9
##bcftoolsCommand=mpileup -O b -o results/bcf/SRR2584866_raw.bcf -f data/ref_genome/ecoli.fa results/bam/SRR2584866.aligned.sorted.bam
##reference=file://data/ref_genome/ecoli.fa
##contig=<ID=NC_000913.2,length=4641652>
##ALT=<ID=*,Description="Represents allele(s) other than observed.">
##INFO=<ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=<ID=IDV,Number=1,Type=Integer,Description="Maximum number of reads supporting an indel">
##INFO=<ID=IMF,Number=1,Type=Float,Description="Maximum fraction of reads supporting an indel">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read depth">
##INFO=<ID=VDB,Number=1,Type=Float,Description="Variant Distance Bias for filtering splice-site artefacts in RNA-seq data (bigger is better)",V
ersion="3">
##INFO=<ID=RPB,Number=1,Type=Float,Description="Mann-Whitney U test of Read Position Bias (bigger is better)">
##INFO=<ID=MQB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality Bias (bigger is better)">
##INFO=<ID=BQB,Number=1,Type=Float,Description="Mann-Whitney U test of Base Quality Bias (bigger is better)">
##INFO=<ID=MQSB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality vs Strand Bias (bigger is better)">
##INFO=<ID=SGB,Number=1,Type=Float,Description="Segregation based metric.">
##INFO=<ID=MQ0F,Number=1,Type=Float,Description="Fraction of MQ0 reads (smaller is better)">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="List of Phred-scaled genotype likelihoods">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=ICB,Number=1,Type=Float,Description="Inbreeding Coefficient Binomial test (bigger is better)">
##INFO=<ID=HOB,Number=1,Type=Float,Description="Bias in the number of HOMs number (smaller is better)">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=DP4,Number=4,Type=Integer,Description="Number of high-quality ref-forward , ref-reverse, alt-forward and alt-reverse bases">
##INFO=<ID=MQ,Number=1,Type=Integer,Description="Average mapping quality">
##bcftools_callVersion=1.9+htslib-1.9
```

```

##FILTER=<ID=PASS,Description="All filters passed">
##bcftoolsVersion=1.9+htslib-1.9
##bcftoolsCommand=mpileup -O b -o results/bcf/SRR2584866_raw.bcf -f data/ref_genome/ecoli.fa results/bam/SRR2584866.aligned.sorted.bam
##reference=file://data/ref_genome/ecoli.fa
##contig=<ID=NC_000913.2,length=4641652>
##ALT=<ID=*,Description="Represents allele(s) other than observed.">
##INFO=<ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=<ID=IDV,Number=1,Type=Integer,Description="Maximum number of reads supporting an indel">
##INFO=<ID=IMF,Number=1,Type=Float,Description="Maximum fraction of reads supporting an indel">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read depth">
##INFO=<ID=VDB,Number=1,Type=Float,Description="Variant Distance Bias for filtering splice-site artefacts in RNA-seq data (bigger is better)",Version="3">
##INFO=<ID=RPB,Number=1,Type=Float,Description="Mann-Whitney U test of Read Position Bias (bigger is better)">
##INFO=<ID=MQB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality Bias (bigger is better)">
##INFO=<ID=BQB,Number=1,Type=Float,Description="Mann-Whitney U test of Base Quality Bias (bigger is better)">
##INFO=<ID=MQSB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality vs Strand Bias (bigger is better)">
##INFO=<ID=SGB,Number=1,Type=Float,Description="Segregation based metric.">
##INFO=<ID=MQ0F,Number=1,Type=Float,Description="Fraction of MQ0 reads (smaller is better)">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="List of Phred-scaled genotype likelihoods">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=ICB,Number=1,Type=Float,Description="Inbreeding Coefficient Binomial test (bigger is better)">
##INFO=<ID=HOB,Number=1,Type=Float,Description="Bias in the number of HOMs number (smaller is better)">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=DP4,Number=4,Type=Integer,Description="Number of high-quality ref-forward , ref-reverse, alt-forward and alt-reverse bases">
##INFO=<ID=MQ,Number=1,Type=Integer,Description="Average mapping quality">
##bcftools_callVersion=1.9+htslib-1.9
##bcftools_callCommand=call --ploidy 1 -m -v -o results/bcf/SRR2584866_variants.vcf results/bcf/SRR2584866_raw.bcf; Date=Tue Sep 1 15:38:47 2020
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT results/bam/SRR2584866.aligned.sorted.bam
NC_000913.2 301 . CTTTTTTTTT CTTTTTTTTT 26.3022 . INDEL;IDV=14;IMF=1;DP=14;VDB=0.462635;SGB=-0.680642;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,3,4;MQ=60
=0;AC=1;AN=1;DP4=0,2,9,3;MQ=60 GT:PL 1:53,0
NC_000913.2 393 . T G 183 . DP=7;VDB=0.108365;SGB=-0.636426;MQSB=1.01283;MQ0F=0;AC=1;AN=1;DP4=0,0,3,4;MQ=60
GT:PL 1:213,0
NC_000913.2 588 . G A 147 . DP=8;VDB=0.0980762;SGB=-0.590765;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,3,2;MQ=60 G
T:PL 1:177,0
NC_000913.2 774 . T C 225 . DP=14;VDB=0.3048;SGB=-0.670168;MQSB=0.916482;MQ0F=0;AC=1;AN=1;DP4=0,0,7,3;MQ=60
GT:PL 1:255,0
NC_000913.2 867 . C T 121 . DP=5;VDB=0.689691;SGB=-0.590765;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,4,1;MQ=60
T:PL 1:151,0
NC_000913.2 939 . G A 174 . DP=8;VDB=0.391417;SGB=-0.636426;MQSB=1;MQ0F=0;AC=1;AN=1;DP4=0,0,2,5;MQ=60 G
T:PL 1:204,0
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ 
```

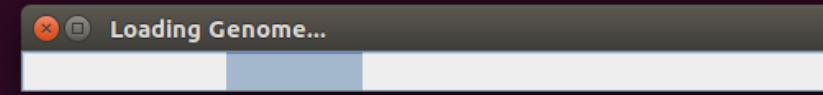
```
 401      411      421      431      441      451      461      471      481      491      501      511      521  
GTGTTGCCGATATTCTGGAAAGCAATGCCAGGCAGGGCAGGTGGCCACCGTCTCTGCCCCCGCCAAAATACCCAACCACCTGGTGGCGATGATTGAAAAAACCATAGCGGCCAGGATGCTTACCCAAATATCAGCGAT  
|G...  
.G.  
.G.  
,g,  
,g,  
,g,  
,g,  
.G. ..
```



Loading Genome...

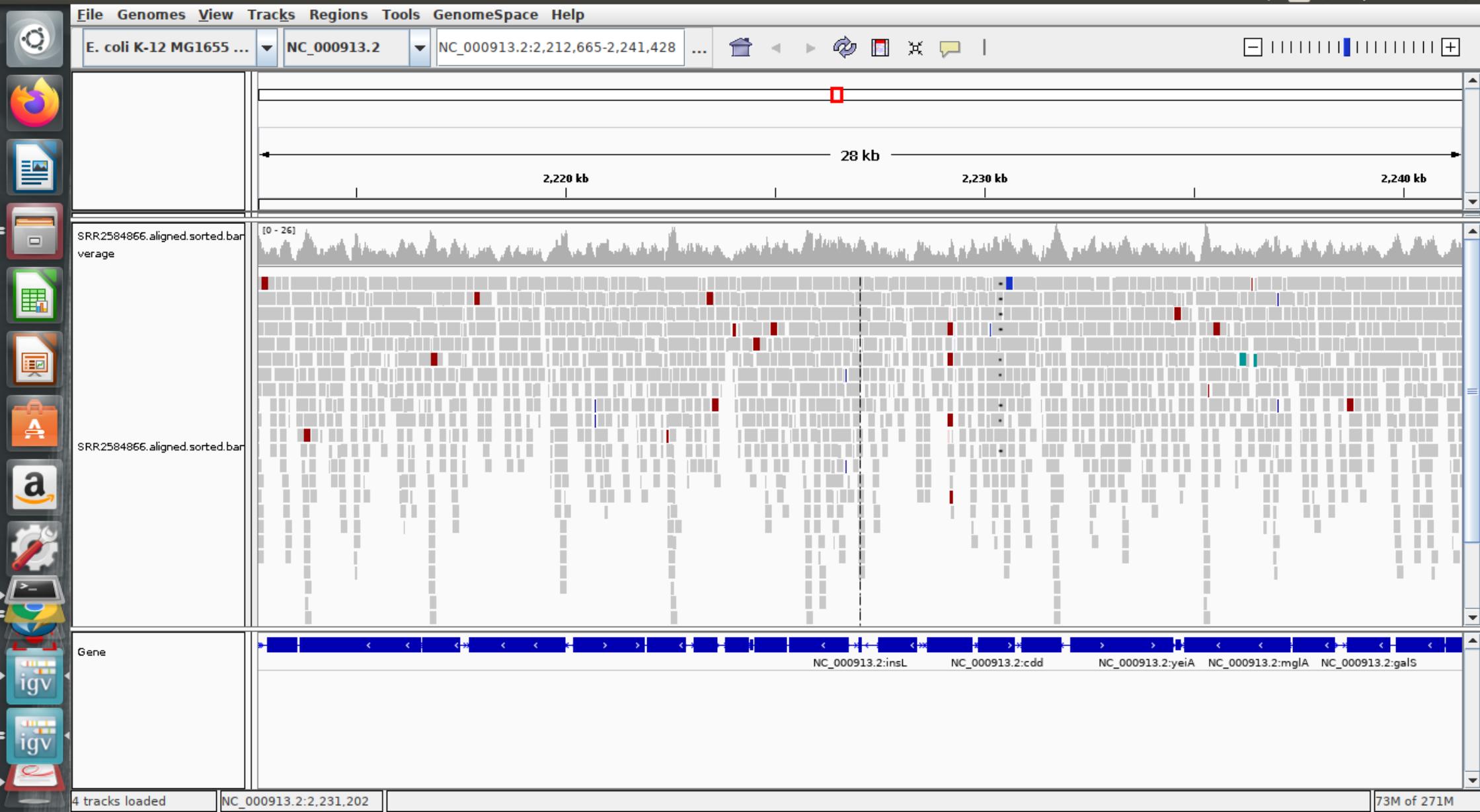
En 3:51 PM

```
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ samtools tview results/bam/SRR2584866.aligned.sorted.bam d  
ata/ref_genome/ecoli.fa  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ^C  
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ igv  
INFO [2020-09-01 15:50:46,905] [DirectoryManager.java:163] [main] IGV Directory: /home/sharmaimr/igv  
INFO [2020-09-01 15:50:46,905] [DirectoryManager.java:163] IGV Directory: /home/sharmaimr/igv  
INFO [2020-09-01 15:50:47,020] [Main.java:99] [main] Startup IGV Version user (0)01/14/2016 11:23 PM  
INFO [2020-09-01 15:50:47,022] [Main.java:100] [main] Java 1.8.0_192  
INFO [2020-09-01 15:50:47,023] [DirectoryManager.java:72] [main] Fetching user directory...  
INFO [2020-09-01 15:50:48,037] [Main.java:101] [main] Default User Directory: /home/sharmaimr  
INFO [2020-09-01 15:50:48,039] [Main.java:102] [main] OS: Linux  
GLib-GIO-Message: 15:50:48.851: Using the 'memory' GSettings backend. Your settings will not be saved or shared with other applications.  
INFO [2020-09-01 15:50:59,997] [GenomeManager.java:145] [main] Loading genome: /home/sharmaimr/igv/genomes/U00096.2.genome
```



sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible: ~/ngs/snp/snp01sep2020/results/sam En 3:54 PM

```
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ samtools tview results.bam/SRR2584866.aligned.sorted.bam
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ^C
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ igv
INFO [2020-09-01 15:50:46,905] [DirectoryManager.java:163] [main] IGV Directory: /home/sharmaimr/igv
INFO [2020-09-01 15:50:46,905] [DirectoryManager.java:163] IGV Directory: /home/sharmaimr/igv
INFO [2020-09-01 15:50:47,020] [Main.java:99] [main] Startup IGV Version user (0)01/14/2016 11:23 PM
INFO [2020-09-01 15:50:47,022] [Main.java:100] [main] Java 1.8.0_192
INFO [2020-09-01 15:50:47,023] [DirectoryManager.java:72] [main] Fetching user directory...
INFO [2020-09-01 15:50:48,037] [Main.java:101] [main] Default User Directory: /home/sharmaimr
INFO [2020-09-01 15:50:48,039] [Main.java:102] [main] OS: Linux
GLib-GIO-Message: 15:50:48.851: Using the 'memory' GSettings backend. Your settings will not be saved or shared with other applications.
INFO [2020-09-01 15:50:59,997] [GenomeManager.java:145] [main] Loading genome: /home/sharmaimr/igv/genomes/U00096.2.genome
INFO [2020-09-01 15:51:05,291] [GenomeManager.java:192] [main] Genome loaded. id= U00096.2
ERROR [2020-09-01 15:51:05,787] [CommandListener.java:124] [Thread-6] java.net.BindException: Address already in use (Bind failed)
INFO [2020-09-01 15:51:11,836] [GenomeManager.java:145] [pool-1-thread-6] Loading genome: https://s3.amazonaws.com/igv.broadinstitute.org/genomes/NC_000913.3.gbk
INFO [2020-09-01 15:51:37,350] [GenomeManager.java:192] [pool-1-thread-6] Genome loaded. id= NC_000913
^CINFO [2020-09-01 15:52:32,379] [ShutdownThread.java:51] [Thread-2] Shutting down
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ ls
data results snpecoli.pdf
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020$ cd results/sam/
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/results/sam$ ls
SRR2584866.aligned.sam
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/results/sam$ head -n5 SRR2584866.aligned.sam
@SQ SN:NC_000913.2 LN:4641652
@PG ID:bwa PN:bwa VN:0.7.17-r1188 CL:bwa mem ecoli.fa trimmed_fastq_small/SRR2584866_1.trim.sub.fastq trimmed_fastq_small/SRR2584866_2.trim.sub.fastq
SRR2584866.7 83 NC_000913.2 3655777 60 26M = 3655152 -651 GCTTCCCATTGCCAATTAAATACATGTG IJJJJJJJJJJJJHHHHHFEEEE
CC@ NM:i:0 MD:Z:26 MC:Z:150M AS:i:26 XS:i:0
SRR2584866.7 163 NC_000913.2 3655152 60 150M = 3655777 651 TCCTCAATTAAAGCGGATCAATGAGCTGGTACGCCATCAGCATATTGATTATCT
GGTGTGAATTTCAGGCTTACGGTAGCTGGCTACGCTGCCACACAGATTAGCTAATTGAAACGCCCTTCACCCCTGCCATACCTTTAAATAATC C@CCCCFFGHGGGGJJJJJJJJIIHIIJJJJJJJJJJJJJJ
IIIJIJJJJJJIIJJHHHHHHHFFFFFDDDDDDDDDDDDDDDBBDDDDDDDDDBDDDDDDDEDEDDDDDDDBDDCDDDDDDBDDDDDDDDDEDED NM:i:0 MD:Z:150 MC:Z:26M A
S:i:150 XS:i:0
SRR2584866.8 99 NC_000913.2 2967216 60 26M = 2967713 647 GCTGATATTCTGCAGCAGTACCGGC @@@DBDDDHFFHIFFGBH<FHC
GDG NM:i:0 MD:Z:26 MC:Z:150M AS:i:26 XS:i:0
(bioinfo) sharmaimr@sharmaimr-HP-Pavilion-x360-Convertible:~/ngs/snp/snp01sep2020/results/sam$
```

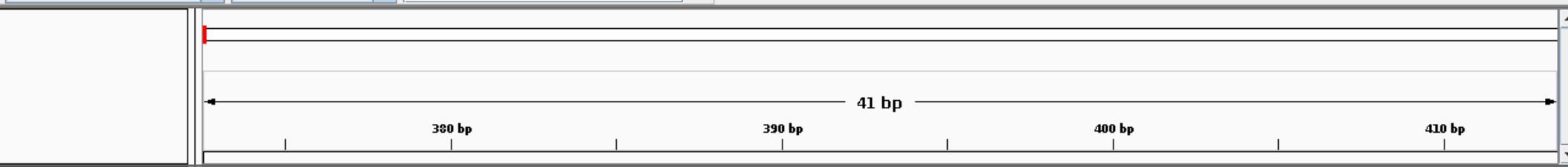
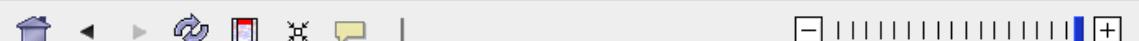


File **Genomes** **Tracks** **Regions** **Tools** **GenomeSpace** **Help**

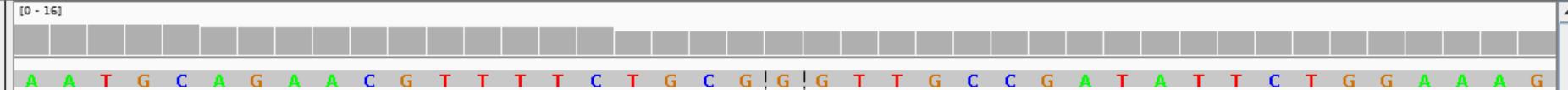
E. coli K-12 MG1655 ...

NC_000913.2

NC_000913.2:373-413



SRR2584866.aligned.sorted.bam
coverage



SRR2584866.aligned.sorted.bam

A A T G C

```
A A T G C A G A A C G T T T T T C
A A T G C A G A A C G T T T T T C T G C G G G T T T G C C C G A T A T T C T G G A A A G
A A T G C A G A A C G T T T T T C T G C G G G T T T G C C C G A T A T T C T G G A A A G
A A T G C A G A A C G T T T T T C T G C G G G T T T G C C C G A T A T T C T G G A A A G
A A T G C A G A A C G T T T T T C T G C G G G T T T G C C C G A T A T T C T G G A A A G
A A T G C A G A A C G T T T T T C T G C G G G T T T G C C C G A T A T T C T G G A A A G
A A T G C A G A A C G T T T T T C T G C G G G T T T G C C C G A T A T T C T G G A A A G
```

Sequence

>>

NC_000913.2:thrA

Gene

File Genomes View Tracks Regions Tools GenomeSpace Help

ecoli.fa NC_000913.2 NC_000913.2:373-413 ...

380 bp 390 bp 400 bp 410 bp 41 bp

SRR2584866.aligned.sorted.bam
verage

[0 - 16]

A A T G C A G A A C G T T T T C T G C G G G T T G C C C G A T A T T C T G G A A A A G

A A T G C

SRR2584866.aligned.sorted.bam

A A T G C A G A A C G T T T T C T G C G G G T T G C C C G A T A T T C T G G A A A A G

A A T G C A G A A C G T T T T C T G C G G G T T G C C C G A T A T T C T G G A A A A G

A A T G C A G A A C G T T T T C T G C G G G T T G C C C G A T A T T C T G G A A A A G

A A T G C A G A A C G T T T T C T G C G G G T T G C C C G A T A T T C T G G A A A A G

A A T G C A G A A C G T T T T C T G C G G G T T G C C C G A T A T T C T G G A A A A G

A A T G C A G A A C G T T T T C T G C G G G T T G C C C G A T A T T C T G G A A A A G

Sequence →

A A T G C A G A A C G T T T C T G C G T G T G C C G A T A T T C T G G A A A A G

N M Q E R F F L R V L C A P D I F S W G E K K

K C R T F S A C V L C P R I Y S F G K K

3 tracks loaded NC_000913.2:399 420M of 1,363M

The image shows the Integrated Genome Viewer (IGV) software interface. At the top, the menu bar includes File, Genomes, View, Tracks, Regions, Tools, GenomeSpace, and Help. The title bar shows 'ecoli.fa' and 'NC_000913.2'. Below the title bar, a navigation bar includes icons for back, forward, search, and other functions. The main window displays a genomic track for 'NC_000913.2:373-413' with four sub-regions labeled 380 bp, 390 bp, 400 bp, and 410 bp. A 41 bp region is highlighted with an orange box. On the left, a sidebar lists 'SRR2584866.aligned.sorted.bam' and 'verage' under the 'bam' category, and 'SRR2584866.aligned.sorted.bam' under the 'fastq' category. The bottom section shows a sequence alignment with a green box highlighting a segment. To the right, a protein sequence is shown with various amino acids (N, M, Q, E, R, F, F, L, R, V, L, C, A, P, D, I, F, S, W, G, E, K, K). The status bar at the bottom indicates '3 tracks loaded' and 'NC_000913.2:399 420M of 1,363M'.