

## Read simulation Using ART

### Ubuntu Installation

```
cd ..  
wget https://github.com/samtools/samtools/releases/download/1.9/samtools-1.9.tar.bz2  
tar -xjf samtools-1.9.tar.bz2  
cd samtools-1.9  
make
```

**sudo apt install art-nextgen-simulation-tools**

### Isolate a chromosome sequence in a new fasta file

```
samtools faidx reference.fa chr1 > chr1.fa
```

```
Art_illumina -ss HS20 -sam -i  
/home/lajoyce/Documents/sequence_alignment/chromosome_2.fasta -l 100 -c 1 -o chr2_read
```

```
CL:art_illumina -ss HS20 -sam -i  
/home/lajoyce/Documents/sequence_alignment/chromosome_2.fasta -l 100 -c 1 -o chr2_read  
-rs 1686537059
```

### How to download bwa-mem2 binaries

```
# Use precompiled binaries (recommended)  
curl -L  
https://github.com/bwa-mem2/bwa-mem2/releases/download/v2.2.1/bwa-mem2-2.2.1_x64-linux  
.tar.bz2 \  
| tar jxf -
```

```
bwa-mem2-2.2.1_x64-linux/bwa-mem2 index ref.fa  
bwa-mem2-2.2.1_x64-linux/bwa-mem2 mem ref.fa read1.fq read2.fq > out.sam
```

### Alignment Using bwa-mem2

```
CL:bwa-mem2-2.2.1_x64-linux/bwa-mem2 mem chromosome_2.fasta chr2_read.fq
```

### Run for 1 million coverage

```
-c --rcount number of reads/read pairs to be generated per sequence(not be used together with  
-f/--fcov) - 1 million  
-l --len the length of reads to be simulated - 100  
-rs --rndSeed the seed for random number generator (default: system time in second) - 1000000
```

HS20 - HiSeq 2000 (100bp)

## Screenshot of run

```
lajoyce@Ubuntu-Age:~/Documents/sequence_alignment/art_bin_MountRainier$ art_illumina -ss HS20 -sam -i /home/lajoyce/Documents/sequence_alignment/chromosome_2.fasta -l 100 -c 1000000 -rs 1000000 -o /home/lajoyce/Documents/sequence_alignment/chromosome/reads_chr2/chr2_reads_1million
o /home/lajoyce/Documents/sequence_alignment/chromosome/reads_chr2/chr2_reads_10_2
=====ART=====
ART_illumina (2008-2016)
Q Version 2.5.8 (June 6, 2016)
Contact: Weichun Huang <whduke@gmail.com>
-----

Single-end Simulation

Total CPU time used: 23.8725

The random seed for the run: 1000000

Parameters used during run
  Read Length:      100
  Genome masking 'N' cutoff frequency: 1 in 100
  Fold Coverage:    0X
  Profile Type:     Combined
  ID Tag:

Quality Profile(s)
  First Read:  HiSeq 2000 Length 100 R1 (built-in profile)

Output files

FASTQ Sequence File:
  /home/lajoyce/Documents/sequence_alignment/chromosome/reads_chr2/chr2_reads_1million.fq

ALN Alignment File:
  /home/lajoyce/Documents/sequence_alignment/chromosome/reads_chr2/chr2_reads_1million.aln

SAM Alignment File:
  /home/lajoyce/Documents/sequence_alignment/chromosome/reads_chr2/chr2_reads_1million.sam
```

## Overview of 1million coverage fastq file

```
lajoyce@Ubuntu-Age: ~/Documents/sequence_alignment/chromosome/reads_chr2
GNU nano 6.2 chr2_reads_1million.fq
@NC_000002.12-1000000
CTTGCAATGCACCCATAGGTATATATCCTAGAGAAATCATGCTTATGTGCTACAGAATATATACGTTCAAGATTGTTTATATATCCAAACT
+
B@CFF;FDHD(HCCHJIHJJJJII(JJ?J>IEJ?JJIDFID)JGGIJJFHF7DJIDHFEHF-ID>HBF(ECBD@C&HDEB&;CBH(DDCE?F(3C<DDD
@NC_000002.12-999999
GCCACACGACATCAACGACCCGATACCAAACTCACCATTCTTGCTAACCTCCTTCCCAACAAAACTGGGGTGACTTCTCCTCTTCTCCTGC
+
CCCF=F@DFHHG):GJJJI:HHFGJI.AJJ(IIIEJGGJIIJJJJJDFJII>JHH'IDJHJG@<CFDCA@>B'8DEHH:FDE(DCDD(EDDD?DDDDCC
@NC_000002.12-999998
CCAGCTTGCTCTGTATTACATAAACTTGAAGAATGAAAGTGCTTGAGGCATAAAAGTATTCATGCCTATTTTAAAGTTTCCATTATGAATAAA
+
@C=F#FF(F(HHHIJJJ())GGJJJEJGHIJJGIGGHFGJE;IDHIJGIJJHFIJHJ)DEIH7@JEDCCFD.JHDECCD,C8CAD(,D+BEDF(+EDC
@NC_000002.12-999997
TATTGACTTGCTCCCTCTCTACTTAGCTAGGCAACTATCTATCTGATCTGAAACCTAGACCTAGAAATAGAAAAGATTATTGACAAATGTTAACAAAA
+
;C@FD=FDHHHH)JGHEJHJI(I@JJJJIIJJJJJJ>GHICBI*H.DGG9JFH.DGI@AD7H>JD0FED>AECHDFIFBCH?DD@95(AACBEDACDE
@NC_000002.12-999996
TTCTCTACAGGCAGCCATCAGCAAGTCTGTGAGCTCTACTCTGAAATACATCCCAAGCCATGACACTTCTCTCTTACACTGCCACTAACCTTGT
+
C:CFFF=FHHHFHI9IJJG@JIJHJIDIJHJIFBCED/JJG9JJ>JJJDIJG;IJDEII(FBCHDEJDCIHCGDCECB+(+ED<DCCF>>FDCACDDC
[ Read 3973220 lines ]
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location  M-U Undo
^X Exit      ^R Read File ^V Replace   ^U Paste     ^J Justify   ^_ Go To Line M-E Redo
```