Project – Sequence Alignment with Bowtie2

Background

- Sequence alignment
  - Compare and align sequence data (reads) to reference to identify regions of similarity (e.g. genes)
- Bowtie2
  - A multi-threaded tool for aligning sequencing reads to long reference sequences

```
Read:       GACTGGGCGATCTCGACTTCG
            |||||   ||||||||||| |||
Reference:  GACTG--CGATCTCGACATCG
```

Tasks

- Run bowtie2 with 1, 2, 4 and 8 threads
  - Record the wall clock time and memory usage
- Use Python (or any method you prefer) to estimate wall clock time and memory usage for the cases with 16 and 32 threads
- Run bowtie2 again with 16 and 32 threads
  - Again, record the wall clock time and memory usage
  - Compare them to the estimated values

Step 1: Locate Files

- `cd` to the directory where all files are located
  `cd /work/<user name>/bowtie2_yeast`

  – `<user name>` should be replace by your training
    account "hpctrnxx"

```
[hpctrn60@mike004 bowtie2_yeast]$ ls -al
total 2069360
drwxr-xr-x 2 hpctrn60 Users        4096 May 28 13:33 .
drwx------ 3 hpctrn60 Users        4096 May 28 13:33 ..
-rw-r--r-- 1 hpctrn60 Users 2080972366 May 28 11:53 read1.fastq
-rw-r--r-- 1 hpctrn60 Users     8247244 May 28 13:27 yeast_ref.1.bt2
-rw-r--r-- 1 hpctrn60 Users     3039284 May 28 13:27 yeast_ref.2.bt2
-rw-r--r-- 1 hpctrn60 Users         161 May 28 13:27 yeast_ref.3.bt2
-rw-r--r-- 1 hpctrn60 Users     3039277 May 28 13:27 yeast_ref.4.bt2
-rw-r--r-- 1 hpctrn60 Users    12400379 May 28 13:27 yeast_ref.fa
-rw-r--r-- 1 hpctrn60 Users     8247244 May 28 13:27 yeast_ref.rev.1.bt2
-rw-r--r-- 1 hpctrn60 Users     3039284 May 28 13:27 yeast_ref.rev.2.bt2
```

Step 2: Run Bowtie2

- ## Run bowtie2 with 1, 2, 4 and 8 threads
  - ## Command:

```
time bowtie2 --threads <number of threads> -x yeast_ref -U read1.fastq -S yeast1.sam
```

  - ## Ex: with 2 threads

```
time bowtie2 --threads 2 -x yeast_ref -U read1.fastq -S yeast1.sam
```

Thread	Time (seconds)	Memory (MB)
1		
2		
4		
8		

Step 3: Estimate

- Use Python (or any method you prefer) to estimate wall clock time and memory usage for the cases with 16 and 32 threads

Thread	Time (seconds)	Memory (MB)
1		
2		
4		
8		
16		
32		

Step 4: Run Bowtie2 Again

- Run bowtie2 again with 16 and 32 threads
- Compare the observed data with your estimates

Thread	Time (seconds)		Memory (MB)	
1				
2				
4				
8				
16	Estimate	Measured	Estimate	Measured
32	Estimate	Measured	Estimate	Measured

Step 5: What Do You Learn?