Dear Awesome PA,

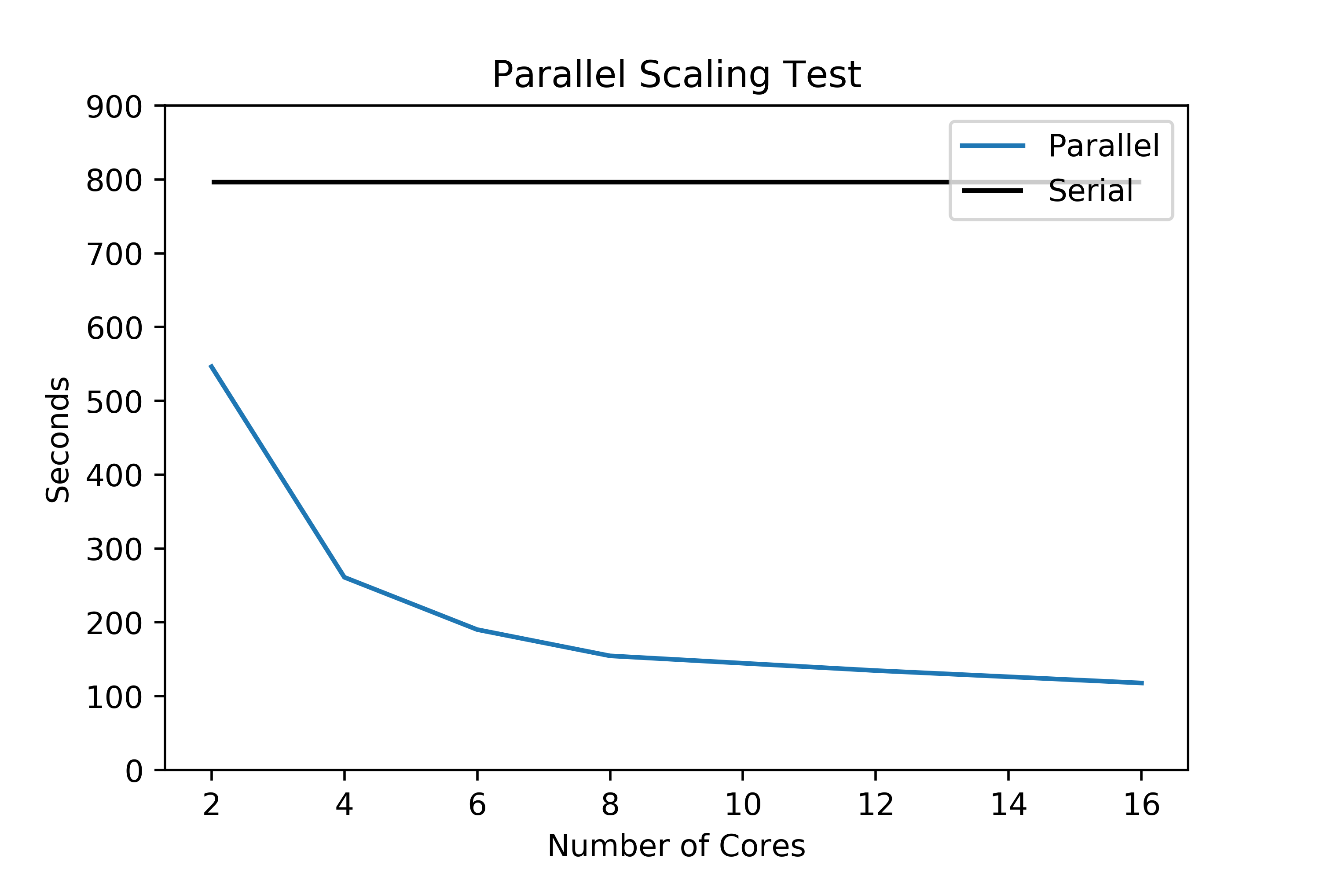
I have gotten your code running in parallel, and fixed issues involving the submission scripts in general to make them less error prone. I have even generated an “example” directory for you so that you can start from a working workflow, with the only changes needed being file names and different downloads. (Though the directory is quite big, even compressed. Make sure you have half a gigabyte of free space before downloading it, and do not run it without at least 5 gigabytes of free space.)

Specifically, you had problems with the requested time and memory, but that can really only be fixed through trying. I also made it so that all commands would be run inside of the directory where the submission script was, to increase versatility of the scripts. I also cleaned up your sruns, essentially you were calling multiple cores with srun in an sbatch submission script that was calling multiple cores. This can cause problems, so make sure you set the number of cores through the sbatch interface. Other than that, there actually were not many problems with your scripts.

The parallel alignment requires you call multiple cores through sbatch and multiple threads through the BWA command, so watch out for that. This parallel section also scaled quite well, and I include a figure to give a rough idea of how it scales so you can choose the number of cores you deem best.

P.S. Samtools cannont process gzipped files, so make sure you decompress files before viewing them with SAMtools.

Research Facilitator



|  |  |
| --- | --- |
| Truncated name of Command | Average Time |
| wget NIST1 | 146.651 SECONDS |
| wget NIST2 | 111.871 SECONDS |
| wget Celegens.toplevel | 20.036 SECONDS |
| index NIST1 | 87.40 minutes |
| index NIST2 | 87.35 minutes |
| index Celegens.toplevel | 1.38 minutes |
| Index Dataset01\_1 | 8.69 minutes |
| Index Dataset01\_2 | 8.90 minutes |
| Serial Assembly (Small Sets) | 13.28 minutes |
| Parallel Assembly (2 cores) | 9.11 minutes |
| Parallel Assembly (4 cores) | 4.35 minutes |
| Parallel Assembly (6 cores) | 3.17 minutes |
| Parallel Assembly (8 cores) | 2.58 minutes |
| Parallel Assembly (12 cores) | 2.24 minutes |
| Parallel Assembly (16 cores) | 1.96 minutes |
| SAMtools prep | 3.99 minutes |