This is a pipeline for variant calling.

1. We did the quality control of our reads using fastqc
2. We trimmed our reads using fastp. We didn't use cutadapt. Indeed with cutadapt, we need to look at the results of fastqc and notice the sequence we will trim (overrepresented sequences). If we did that, our script won't be automatic (run it once). With fastp, we won't need to run the script twice.
3. Then we use BWA to index our reference genome and to align the trimmed reads to that reference.
4. We got a .sam file as a result of step 3. We will convert that file to .bam using samtools. We also use it to sort the bam file.
5. We use freebayes for variant calling. As a result we got .vcf file

This pipeline is run once and takes three arguments.

1. The first one is the directory containing our reads. The pipeline is able to take many reads. The only thing is to make sure that reverse and forward are labeled as follow:
   reverse: *r1*.fastqc or r1*.fastqc.gz
   forward: *r2*.fastqc or r2*.fastqc.gz
   *=any string
2. The second argument is the directory containing the reference genome. That latter can be .fa or .fa.gz
3. The third one will be a directory to output the results.

Warnings: The script will create a directory named **output** and **results**; be careful to run it in a directory that doesn't contain these two names.

bash pipeline.sh 1st arg 2nd arg 3rd arg