

All changes below are specific to macOS and even more specifically to iMac (late 2009 model, 3.06 GHz Intel Core 2 Duo, 6GB 1067 DDR memory) macOS High Sierra v 10.13.6. One of the most frustrating issues I had was being constrained to this version of macOS and this model computer. The computer is too old to support newer updates and it seems this version of High Sierra had lots of issues with bioinformatics programs. That said, Xcode and CommandLineTools of particular versions were installed and found to be the best options.

1. Compiler updates, fixes, Xcode (specific version), Command line tools (specific version)

- Xcode: v9.4.1
- CommandLineTools (CLT): CommandLineTools (macOS 10.13) for Xcode 9.4.1 (<https://developer.apple.com/download/more/>)

These versions of Xcode and CLT required certain libraries to be installed. There was a problem where installation of different CLT was overwriting and not including certain header files such as zlib.h so the workaround was to go to /usr/local/include and change it to usr/local/include_old so as to have installation of CLT install a new /usr/local/include/ directory full of all header files including the correct ones. This mostly had to deal with C++ compiler problems and the reference is:

<https://stackoverflow.com/questions/46872922/broken-c-std-libraries-on-macos-high-sierra-10-13>.

2. gsl and RcppGSI errors: R version: 3.5

- used this site (<https://thecoatlessprofessor.com/programming/r-compiler-tools-for-rcpp-on-macos-before-r-3.6.0/>) for help with compiler errors between R and C++.
- .bash_profile:
#to use clang compiler
export PATH=/usr/local/clang6/bin:\$PATH
export PATH=/usr/local/clang4/bin/clang++:\$PATH
- .Renvron file:
PATH="/usr/local/clang6/bin:\${PATH}"
PATH="/usr/local/clang4/bin/c++:\${PATH}"
PATH="/usr/local/clang6/bin:\${PATH}"

3. Added haplogrep script Line 300

- (java -jar /Users/Snow_Lab_bioinformatics/02_Programs/haplogrep/haplogrep-2.1.20.jar --in \${NAME}.q30.chrM.rescaled.mia.consensus.fasta --format fasta --out \${NAME}.consensus.haplogrep.txt #adding code for haplogrep analysis 6/24/19 TB)

4. Edited command for variant call line

- Line 254 and 257, changed to bcftools since samtools is deprecated. Bcftools mpileup didn't change much from previous samtools mpileup command. Using bcftools call as suggested in <https://samtools.github.io/bcftools/howtos/variant-calling.html>. Command is now: bcftools call --ploidy 1 -mv -Ov -o name.jv.vcf

5. Other/Miscellaneous changes

- Edited one of the columns in the final output summary file % Mapped (Mapped reads/% kept)
- Using this sequence for adapters
AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC
- Changed sort to srt throughout script. Found later on in the script that something was listed as only srt and messed up with running script so just changed all to srt.
- Changed pmd_hist_V3.R #1 to `pmd<-read.table(paste(args[1],".pmdscores.txt", sep=""), header=F)` as suggested by reddit user RWeThereYet17 (https://www.reddit.com/r/rprogramming/comments/bzgcil/readtables_error_in_r/)
- warning: text-based stub file TBD/file/path and library file library/file/path are out of sync. Falling back to library file for linking: Solution-> forcibly remove and reinstall command line tools (<http://sd.jtimothyking.com/2018/07/26/stub-file-and-library-file-out-of-sync/>)
- Bamqc will fail if re-running analysis in same directory structure with previous bams in sub-directories.

Line 83, changed -o option to 10, added -S for spinner because I found that I couldn't get it to run without this option

Line 135, added > to output as it wouldn't work properly without it

Line 300, added haplogrep, outputs into intermediate files folder

Lines 392-403, needed to re-write as R CMD BATCH to invoke command

I had to make other changes not associated with this pipeline but to dependency programs, library scripts, etc. that weren't cooperating well between different versions of things since I was limited to my ability to update software and hardware.