**UNEAK and post-processing step by step**

1. Run UNEAK pipeline. Remove BLANK from resulting file, or from key file directly.

2. Run fasFilter.pl to get Query and Hit sequences

Change directory to hapMap folder. At command prompt ‘>’ type

**> perl fasFilter.pl HapMap.fas.txt out.txt out1.txt**

fasFilter.pl must be in the same folder as HapMap.fas.txt

out.txt contains the ‘Query’ and ‘Hit’ columns – copy and paste and add to HapMap.hmc.txt in

columns 2 and 3.

Go into HapMap.hmp.txt – copy column ‘alleles’ to HapMap.hmc.txt column 4

3. At this point it may help to sort the file horizontally to get individuals in order

4. Run PairDuplicates.pl to sort out reverse complement issues

**> perl PairDuplicates.pl HapMap.hmc.txt 281 HapMapPaired.txt**

Where 281 is the sample size and HapMapPaired.txt is the new outfile

5. Run UNEAK\_filter1.pl to call genotypes and filter by observed heterozygosity and number of missing data

**> perl UNEAK\_filter1.pl HapMapPaired.txt 281 HapMapPairedFilt.txt 56 0.75**

Where 56 is the acceptable number of missing data per locus (20% from n=281) and 0.75 is the observed heterozygosity threshold (loci with observed heterozygosity higher than this assumed to be paralogs and are removed).

6. Run MOD\_TO\_GENEPOP\_rrv.pl to create genepop file

**> perl MOD\_TO\_GENEPOP\_rrv.pl HapMapPairedFilt.txt 281 HapMapGENEPOP.txt**

7. Run blastStacksREFSEQrna.pl to blast against mammalian rna REFSEQ database

**> perl blastStacksREFSEQrna.pl HapMapPairedFilt.txt intfile.txt finout.txt**

Where intfile.txt is an intermediate file and finout is the final output. Name these so they don’t

conflict with other file names. Run this on the CBSU workstations. When finished copy and paste

finout.txt into HapMapPairedFilt.txt

Note: some perl scripts were modified by Tom White to meet any data (added the suffix \_rrv).