**Description of the simulated sequence data**

The simulated sequence data can simply be regenerated using QMSim (the 2013 Linux version that is available from: <http://www.aps.uoguelph.ca/~msargol/qmsim/> ) and the parameter file “sequence.txt” using the seed file “seed\_used.prv”.

Subsequently, the program “recode\_genotypes.f90” can be used to generate input files for SNPrune, considering either allele counts (“SNP\_genotypes.txt”) or phased alleles (“phased\_alleles.txt”).

The file “SNP\_genotypes.txt” contains allele counts (0,1,2) for 10,812,225 segregating SNPs for 2500 individuals. The format of the file looks like (the following snap-shot):

50201 01212211001021210112111101011111112221110002020111021

50202 01212211001021210111111101011111111221110002120111021

50203 01212211001021210111111101011111111221110002120111021

50204 01212211001021200112111101011211112221110002020111121

50205 02202220000020200021000000020220001220200002120220020

50206 01112211011021200112111101011211112221110012020111021

50207 12201210111111201122100110120221102111101112111110011

50208 01212211001021210111111101011111111221110002120111021

50209 02202220000020200021000000020220001220200002120220020

50210 11211201112112211212211211111012212112011111111001012

The file “phased\_alleles.txt” contains alleles (0,1) for 10,812,225 segregating SNPs for 2500 individuals. The format of the file looks like (the following snap-shot):

50201 01101110000010100011000000010110001110100001010110010

50201 00111101001011110101111101001001111111010001010001011

50202 00111101001011110101111101001001111111010001010001011

50202 01101110000010100010000000010110000110100001110110010

50203 00111101001011110101111101001001111111010001010001011

50203 01101110000010100010000000010110000110100001110110010

50204 01101110000010100011000000010110001110100001010110010

50204 00111101001011100101111101001101111111010001010001111

50205 01101110000010100011000000010110001110100001010110010

50205 01101110000010100010000000010110000110100001110110010

The input files for PLINK are generated using PLINK and the following command:

plink --bfile SNP\_genotypes.txt

Which generates the files:

* SNP\_genotypes.txt.bed
* SNP\_genotypes.txt.bim
* SNP\_genotypes.txt.fam

The command line argument used to run PLINK for pruning using was of the form (using e.g. a window of 5000 SNPs):

plink --bed SNP\_genotypes.txt.bed --indep-pairwise 5000 500 0.99

And when using the maximum likelihood phasing information it was:

plink --bed SNP\_genotypes.txt.bed --indep-pairphase 5000 500 0.99