GENOME: Whole Genome Coalescent Simulator (2Mb region) Haploview Plot

N=10000 diploid individuals, n=200, nPOP=1, fragment=20000, length=100, numChr=1, #SNP=Poisson (result=4943), rec=1e-6, mut=1e-8
Hudson’s ms (2Mb region, equivalent settings) Haploview Plot

ms 200 -t 800 -r 800 20000, 4Nr=800, 4Nu=800, fragment=20000, result=4658 SNPs

Marker 1-200

Marker 1-500

Marker 1-1000
**GENOME: Whole Genome Coalescent Simulator** (SNPs with MAF > 0.05, 2597 common SNPs) **Haploview Plot**

N=10000 diploid individuals, n=200, nPOP=1, fragment=20000, length=100, numChr=1, #SNP=Poisson (result=4943), rec=1e-6, mut=1e-8
Hudson's ms (SNPs with MAF > 0.05, 2426 common SNPs) Haploview Plot

ms 200 1 -t 800 -r 800 20000, 4Nr=800, 4Nu=800, fragment=20000, result=4658 SNPs
Allele frequency distribution (GENOME)

Allele frequency distribution (Hudson's ms)
Distance is defined as the number of intervening SNPs.
Distribution of LD by genetic distance

![Graph showing mean of r^2 vs. scaled recombination rate](image)

![Graph showing standard deviation of r^2 vs. scaled recombination rate](image)
30 Chromosomes, 2000 unlinked loci (each 1kb), population size = 10,000

Chi-Square statistics of goodness-of-fit test = 29.808
P-Value = 0.372 (for degrees of freedom = 28)
Run Time Comparison

Settings:

Effective population size, $N=10000$ diploid individuals (for 2 pops, $N$=the size of each subpopulation)
1200 chromosomes (600/600 for 2 populations)
15000 fragments for 150Mb region.
Migration rate = $2.5 \times 10^{-4}$ per generation, \((4N_m=10)\)
Mutation rate = $10^{-8}$ per base pair \((4N_u=60000\) for 150 Mb)
Recombination rate = $10^{-8}$ per base pair \((4N_r=60000\) for 150Mb)

<table>
<thead>
<tr>
<th>Length of Region</th>
<th>Number of Populations</th>
<th>GENOME (seconds)</th>
<th>ms (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>150Mb</td>
<td>1</td>
<td>1556 (25.9 mins)</td>
<td>13416 (3.7 hrs)</td>
</tr>
<tr>
<td>150Mb</td>
<td>2</td>
<td>3964 (66.1 mins)</td>
<td>45138 (12.5 hrs)</td>
</tr>
</tbody>
</table>

We simulate 300 sequences of 20Mb, 30Mb or 40Mb using COSI v1.1 and GENOME. COSI crashes when simulating 50Mb and 300 sequences or for longer region and more sequences. Observed that COSI took about 2.8G memory when simulating 30Mb region.

<table>
<thead>
<tr>
<th>Length of Region</th>
<th>COSI v1.1 (seconds)</th>
<th>GENOME (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40M</td>
<td>3089</td>
<td>294</td>
</tr>
<tr>
<td>30M</td>
<td>1658</td>
<td>240</td>
</tr>
<tr>
<td>20M</td>
<td>677</td>
<td>150</td>
</tr>
</tbody>
</table>