

snpparrays_benchmark

June 9, 2016

1 SnpArrayBM - SnpArray Bench Mark

Instead of storing genotype data as two BitArrays, SnpArrayBM stores genotype data as a 3-D (2 x number of people x number of SNPs) BitArray.

```
type SnpArrayBM{N} <: AbstractArray{NTuple{2, Bool}, N}
    A::BitArray{N}
end
```

1.1 Constructor - load from PLINK file

```
In [15]: include("../src/SnpArrays.jl");
         using SnpArrays, BaseTestNext;
         @time sa = SnpArray("../docs/hapmap3");
         @time sabm = SnpArrayBM("../docs/hapmap3");

0.075597 seconds (186 allocations: 1.102 MB)
0.065653 seconds (178 allocations: 1.101 MB)
```

1.2 Test getindex

```
In [2]: n,p = size(sa)
         for j=1:p
             for i=1:n
                 @test sa[i,j] == sabm[i,j];
             end
         end
```

1.3 Test summarize

```
In [14]: @time maf1, _, _, _ = summarize(sa);
         @time maf2, _, _, _ = summarize(sabm);
         for i=1:n
             @test maf1[i] == maf2[i];
         end
```

```
0.072515 seconds (18 allocations: 222.516 KB)
0.057378 seconds (18 allocations: 222.516 KB)
```

1.4 Test GRM

```
In [31]: @time grm(sa; method=:GRM);
          @time grm(sabm; method=:GRM);

0.118506 seconds (15 allocations: 35.231 MB, 0.73% gc time)
0.117168 seconds (15 allocations: 35.231 MB, 0.59% gc time)
```

1.5 Test MoM

```
In [8]: @time grm(sa; method=:MoM);
          @time grm(sabm; method=:MoM);

0.210271 seconds (25 allocations: 35.448 MB, 0.61% gc time)
0.194818 seconds (25 allocations: 35.448 MB, 0.75% gc time)
```

1.6 Test PCA

```
In [43]: @time sc1, _, _ = pca(sa);
          @time sc2, _, _ = pca(sabm);
          for i=1:6
              @show sc1[i], sc2[i]
          end

1.463894 seconds (4.53 k allocations: 110.433 MB, 0.26% gc time)
1.207079 seconds (4.20 k allocations: 102.594 MB, 0.27% gc time)
(sc1[i],sc2[i]) = (-38.74978599130724,38.678595843276966)
(sc1[i],sc2[i]) = (-32.6031247407157,32.68087907528479)
(sc1[i],sc2[i]) = (-22.928752533722,22.954643274071504)
(sc1[i],sc2[i]) = (-35.74279266844716,35.7462233956027)
(sc1[i],sc2[i]) = (-37.17407711393033,37.195669565547675)
(sc1[i],sc2[i]) = (-34.94035819009948,35.00983863987926)
```