snparrays_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 [13]: include("../src/SnpArrays.jl");
         using SnpArrays, BaseTestNext;
         @time sa = SnpArray("../docs/hapmap3");
         @time sabm = SnpArrayBM("../docs/hapmap3");

  0.083832 seconds (186 allocations: 1.102 MB)
  0.061803 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 [12]: @time maf1, _, _, _ = summarize(sa);
         @time maf2, _, _, _ = summarize(sabm);
         for i=1:n
             @test maf1[i] == maf2[i];
         end
```

```
  0.064023 seconds (18 allocations: 222.516 KB)
  0.056310 seconds (18 allocations: 222.516 KB)
```

1.4 Test GRM

```
In [11]: @time grm(sa; method=:GRM);
         @time grm(sabm; method=:GRM);

  0.124313 seconds (15 allocations: 35.231 MB, 0.97% gc time)
  0.138014 seconds (15 allocations: 35.231 MB, 0.75% gc time)
```

1.5 Test MoM

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

  0.228014 seconds (25 allocations: 35.448 MB, 1.41% gc time)
  0.209778 seconds (25 allocations: 35.448 MB, 0.65% gc time)
```

1.6 Test PCA

```
In [7]: @time _, _, v1 = pca(sa);
        @time _, _, v2 = pca(sabm);
        for i=1:6
            @show v1[i], v2[i]
        end

  1.522002 seconds (4.53 k allocations: 110.433 MB, 0.49% gc time)
  1.341803 seconds (4.37 k allocations: 106.513 MB, 0.29% gc time)
(v1[i],v2[i]) = (1841.212290037642,1841.1394804972747)
(v1[i],v2[i]) = (225.36897175041923,225.313378902105)
(v1[i],v2[i]) = (70.73259633894456,70.66751799295083)
(v1[i],v2[i]) = (70.00321616359496,69.87869876413833)
(v1[i],v2[i]) = (69.01676505755543,69.0603365100439)
(v1[i],v2[i]) = (67.85534745015931,67.93095431461947)
```

```
In [ ]:
```