

Paralelizando plink

Ver test 1 aqui: [Plink SET1xSET2](#)

Pensando

1. Una idea es extraer del archivo de sets los marcadores que no esten en la base de datos.
2. La otra es separar el archivo de sets en varios archivos y correr *plink* con SET1xSET2 en lugar de SET1xSET1. Esto permite paralelizar los procesos en varios procesadores con *parallel*

Ejemplo de extraccion

```
grpseries.pl setsTEST5.txt /home/data/Variomics/ADMURimpQC2.bim >
setsTEST5.txt.grp
```

[grpseries.pl](#)

```
#!/usr/bin/perl

use strict; use warnings;
use File::Slurp qw(read_file);
use Array::Utils qw(intersect);

my $sfile = shift;
my $dbfile = shift;

my @series = map {/^(rs\d{1,18})$/} read_file $sfile;
my @dbpairs = map {/\s+(rs\d{1,18})\s+/} read_file $dbfile;

my @chosen = intersect @series, @dbpairs;

foreach my $p (@chosen){
    print "$p\n";
}
```

From:
<https://imagen.fundacioace.com/wiki/> - **Detritus Wiki**

Permanent link:
https://imagen.fundacioace.com/wiki/doku.php?id=genetica:plink_parallel

Last update: **2020/08/04 10:58**

