Analyse SNP data with plink

p-link --tfile rice_ngs --indep-pairwise 50(window size) 3 (step shift) 0.08 --missing-genotype N

p-link --tfile rice_ngs --indep-pairwise 50 3 0.8 --missing-genotype N

Thus, sliding windows of 50, 500, 5000, 50,000, 500,000, or 5,000,000 SNPs were used for the sequence data, that were shifted forward in steps of 10% of the window size, i.e. with 5, 50

how to navigate in different directory

Windows Command Line Tutorial - 1 - Introduction to the Command Prompt

https://www.youtube.com/watch?v=MBBWVgE0ewk

Windows Command Line Tutorial - 2 - Listing Files and Directories

https://www.youtube.com/watch?v=7ABkcHLdG_A

First, you have plink in your disk, maybe document,

Second, you locate this directory in your window command prompt

Then you can run,

Or you can use gPLINK.jar application

Data management tools

PLINK provides a simple interface for recoding, reordering, merging, flipping DNA-strand and extracting subsets of data.

http://zzz.bwh.harvard.edu/plink/dataman.shtml

Determining the best LD Pruning options

http://blog.goldenhelix.com/dkammeraad/determining-best-ld-pruning-options/

Thanks for your request!

A member of our team will be in touch with you shortly.

If you do not hear from us within 24 hours, there could have been an issue processing your request. Should this be the case, please email us at info@goldenhelix.com.
http://goldenhelix.com/forms/eval-request.html

gPLINK

`gPLINK` is a freely-available, Java-based software package that:

- is a GUI that allows construction of many common `PLINK` operations
- provides a simple project management tool and analysis log
- allows for data and computation to be on a separate server (via SSH)
- facilitates integration with Haploview

http://zzz.bwh.harvard.edu/plink/gplink.shtml

Integration with Haploview

http://zzz.bwh.harvard.edu/plink/gplink.shtml#down

Make binary PED file:

Page 24

Working with the binary PED file

Page 25

Order of major operations in PLINK

Page 289

• Perform LD-based pruning of SNP (--indep, --indep-pairwise), then QUIT

p-link --tfile rice_ngs --indep-pairwise 50 3 0.8 --missing-genotype N

292

Generating frequencies for these SNPs,

plink --file test –freq

46

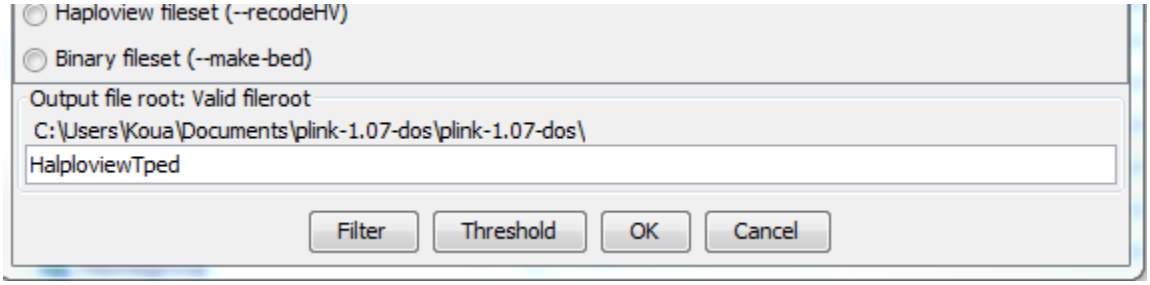
3.4 Transposed filesets

47

The default missing genotype character can be changed with the --missing-genotype option, for example:

plink --file mydata --missing-genotype N

in gPLINK the output file should be with no space



3.3.1 Chromosome codes

The autosomes should be coded 1 through 22. The following other codes can be used to specify other chromosome types:

X X chromosome -> 23

Y Y chromosome -> 24

XY Pseudo-autosomal region of X -> 25

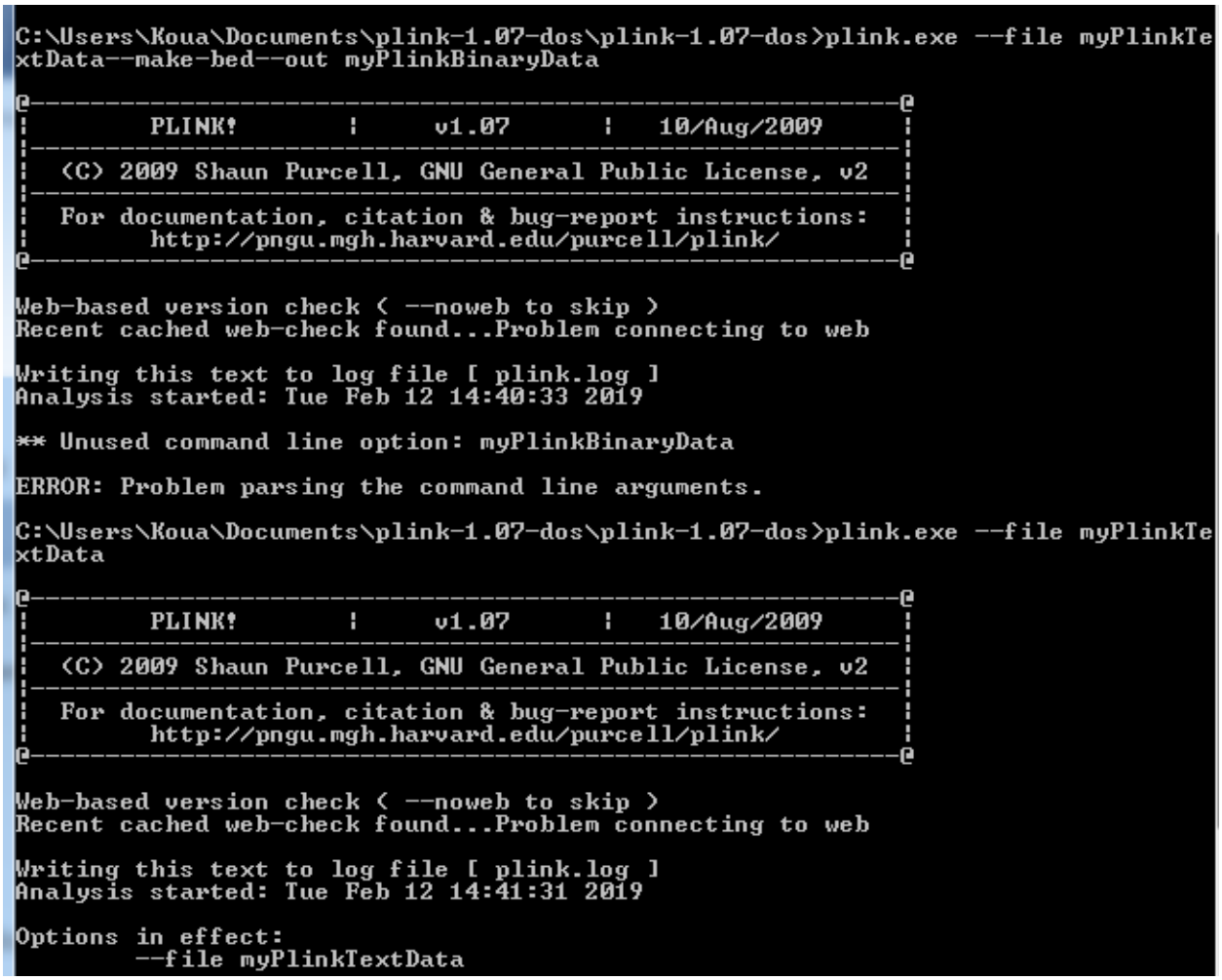
MT Mitochondrial -> 26

45

To subsequently load a binary file, just use --bfile instead of --file

plink --bfile mydata

51



Run all the Briwecs Pop

Briwecs213

Content of the log last run

```
Web-based version check ( --noweb to skip )
Recent cached web-check found...Problem connecting to web

Writing this text to log file [ plinkbriwecsrun1.log ]
Analysis started: Fri Feb 22 12:03:33 2019

** Unused command line option: --missing-genotypes
** Unused command line option: N

ERROR: Problem parsing the command line arguments.

C:\Users\Koua\Documents\plink-1.07-dos>plink.exe --tfile BriwecsNew --indep-pair
wise 3500 350 0.8 --out plinkbriwecsrun1 --missing-genotype N

@----------------------------------------------------------@
|        PLINK!       |     v1.07      |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found...Problem connecting to web

Writing this text to log file [ plinkbriwecsrun1.log ]
Analysis started: Fri Feb 22 12:04:19 2019

Options in effect:
        --tfile BriwecsNew
        --indep-pairwise 3500 350 0.8
        --out plinkbriwecsrun1
        --missing-genotype N


ERROR: No file [ BriwecsNew.tped ] exists.

C:\Users\Koua\Documents\plink-1.07-dos>plink.exe --tfile Briwecs213New --indep-p
airwise 3500 350 0.8 --out plinkbriwecsrun1 --missing-genotype N

@----------------------------------------------------------@
|        PLINK!       |     v1.07      |   10/Aug/2009     |
|----------------------------------------------------------|
|  (C) 2009 Shaun Purcell, GNU General Public License, v2  |
|----------------------------------------------------------|
|  For documentation, citation & bug-report instructions:  |
|        http://pngu.mgh.harvard.edu/purcell/plink/        |
@----------------------------------------------------------@

Web-based version check ( --noweb to skip )
Recent cached web-check found...Problem connecting to web

Writing this text to log file [ plinkbriwecsrun1.log ]
Analysis started: Fri Feb 22 12:05:01 2019

Options in effect:
        --tfile Briwecs213New
        --indep-pairwise 3500 350 0.8
        --out plinkbriwecsrun1
        --missing-genotype N

Reading pedigree information from [ Briwecs213New.tfam ]
213 individuals read from [ Briwecs213New.tfam ]
0 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
0 cases, 0 controls and 213 missing
0 males, 0 females, and 213 of unspecified sex
Warning, found 213 individuals with ambiguous sex codes

Writing list of these individuals to [ plinkbriwecsrun1.nosex ]
42340 (of 42340) markers to be included from [ Briwecs213New.tped ]
Before frequency and genotyping pruning, there are 42340 SNPs
213 founders and 0 non-founders found
Total genotyping rate in remaining individuals is 0.988082
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 42340 SNPs
After filtering, 0 cases, 0 controls and 213 missing
After filtering, 0 males, 0 females, and 213 of unspecified sex
Performing LD-based pruning...
Writing pruned-in SNPs to [ plinkbriwecsrun1.prune.in ]
Writing pruned-out SNPs to [ plinkbriwecsrun1.prune.out ]
Scanning from chromosome 1 to 21

Scan region on chromosome 1 from [ W135_FSCP ] to [ W135_EFEV ]
Pruning SNPs 1 to 2422 of 2422
For chromosome 1, 1628 SNPs pruned out, 794 remaining
Scan region on chromosome 2 from [ W135_GJDU ] to [ W15_AOQJ ]
Pruning SNPs 1 to 2993 of 2993
For chromosome 2, 2081 SNPs pruned out, 912 remaining
Scan region on chromosome 3 from [ W15_AJFB ] to [ W135_EIJO ]
Pruning SNPs 1 to 972 of 972
```

Options in effect:

--tfile Briwecs213New

--indep-pairwise 3500 350 0.8

--out plinkbriwecsrun1

--missing-genotype N

After pruning we obtained **13527 markers** pruned in!

And 28796 markers pruned out

Pruning script well explained

To give a concrete example: the command above that specifies 50 5 0.5 would a) consider a window of

50 SNPs, b) calculate LD between each pair of SNPs in the window, b) remove one of a pair of SNPs if the

LD is greater than 0.5, c) shift the window 5 SNPs forward and repeat the procedure.

To make a new, pruned file, then use something like (in this example, we also convert the standard PED

fileset to a binary one):

plink --file data --extract plink.prune.in --make-bed --out pruneddata