Package 'BEDMatrix'

September 6, 2024

Version 2.0.4

License MIT + file LICENSE

Title Extract Genotypes from a PLINK .bed File

Description A matrix-like data structure that allows for efficient,
convenient, and scalable subsetting of binary genotype/phenotype files
generated by PLINK (<https://www.cog-genomics.org/plink2>), the whole
genome association analysis toolset, without loading the entire file into
memory.

URL https://github.com/QuantGen/BEDMatrix

BugReports https://github.com/QuantGen/BEDMatrix/issues

Depends R (>= 3.4.0)

Imports methods, crochet (>= 2.3.0)

Suggests data.table, LinkedMatrix, tinytest

NeedsCompilation yes

Author Alexander Grueneberg [aut, cre],
Gustavo de los Campos [ctb]

Maintainer Alexander Grueneberg <cran@agrueneberg.info>

Repository CRAN

Date/Publication 2024-09-06 19:20:14 UTC

Contents

1

BEDMatrix-package   *A Package to Extract Genotypes from a PLINK .bed File*

Description

The `BEDMatrix` package provides a matrix-like wrapper around .bed files, one of the genotype/phenotype file formats of PLINK, the whole genome association analysis toolset. `BEDMatrix` objects are created by simply providing the path to a .bed file and once created, they behave similarly to regular matrices with the advantage that genotypes are retrieved on demand without loading the entire file into memory. This allows handling of very large files with limited use of memory.

.bed Files

.bed files (sometimes referred to as binary .ped files) are binary representations of genotype calls at biallelic variants. This very compact file format (2 bits per genotype call) is used and generated by PLINK. .bed files should not be confused with the UCSC Genome Browser's BED format, which is totally different.

A .bed file can be created from a .ped file with PLINK using `plink --file myfile --make-bed`.

See Also

`BEDMatrix-class` to learn more about the `BEDMatrix` class.

BEDMatrix   *Create a BEDMatrix Object from a PLINK .bed File*

Description

This function constructs a new `BEDMatrix` object by mapping the specified PLINK .bed file into memory.

Usage

```
BEDMatrix(path, n = NULL, p = NULL, simple_names = FALSE)
```

Arguments

path	Path to the .bed file (with or without extension).
n	The number of samples. If `NULL` (the default), this number will be determined from the accompanying .fam file (of the same name as the .bed file). If a positive integer, the .fam file is not read and `rownames` will be set to `NULL` and have to be provided manually.

p   The number of variants. If NULL (the default) the number of variants will be
                       determined from the accompanying .bim file (of the same name as the .bed file).
                       If a positive integer, the .bim file is not read and colnames will be set to NULL
                       and have to be provided manually.

simple_names   Whether to simplify the format of the dimension names. If FALSE (the default),
                       row names are concatenations of family IDs, _, and within-family IDs, while
                       column names are concatenations of variant names, _, and allele codes for the
                       first allele of the .bim file (A1). If TRUE, row names are within-family IDs only
                       and column names are variant names only.

Details

.bed files must be accompanied by .fam and .bim files: .fam files contain sample information, and
.bim files contain variant information. If the name of the .bed file is *plink*.bed then the names of
the .fam and .bim files have to be *plink*.fam and *plink*.bim, respectively. The .fam and .bim files are
used to extract the number and names of samples and variants.

For very large .bed files, reading the .fam and .bim files can take a long time. If n and p are provided,
these files are not read and dimnames have to be provided manually.

Currently, only the variant-major mode of .bed files is supported. PLINK2 "dropped" support for
the sample-major mode by automatically converting files in this format to the variant-major mode.
Therefore, it is recommended to run files in sample-major mode through PLINK2 first.

Value

A BEDMatrix object.

See Also

BEDMatrix-package to learn more about the BEDMatrix package, BEDMatrix-class to learn more
about the BEDMatrix class.

Examples

```
# Get the path to the example .bed file
path <- system.file("extdata", "example.bed",
                    package = "BEDMatrix")

# Create a BEDMatrix object the example .bed file
m1 <- BEDMatrix(path)

# Create a BEDMatrix object the example .bed file without loading
# the .fam and .bim files
m2 <- BEDMatrix(path, n = 50, p = 1000)
```

BEDMatrix-class *A Class to Extract Genotypes from a PLINK .bed File*

Description

BEDMatrix is a class that maps a PLINK .bed file into memory and behaves similarly to a regular matrix by implementing key methods such as [, dim, and dimnames. Subsets are extracted directly and on-demand from the .bed file without loading the entire file into memory.

Details

The subsets extracted from a BEDMatrix object are coded as the allelic dosages of the first allele in the .bim file (A1), similarly to .raw files generated with the --recode A argument in PLINK).

Internally, this class is an S4 class with the following slots that should not be relied upon in actual code: xptr, dims, dnames, and path. The .bed file is mapped into memory using mmap on Unix and MapViewOfFile on Windows.

Slots

xptr: An external pointer to the underlying C code.

dims: An integer vector specifying the number of samples and variants as determined by the the accompanying .fam and .bim files or by the n and p parameters of the BEDMatrix constructor function.

dnames: A list describing the row names and column names of the object as determined by the accompanying .fam and .bim files, or NULL if the n and p parameters of the BEDMatrix constructor function were provided.

path: A character string containing the path to the .bed file.

Methods

[: Extract parts of an object

dim: Retrieve the dimension of an object

dimnames: Retrieve the dimnames of an object

dimnames<-: Set the dimnames of an object

as.matrix: Turn the object into a matrix

is.matrix: Test if the object is a matrix

length: Get the length of an object

str: Display the internal structure of an object

show: Display the object

See Also

BEDMatrix to create a BEDMatrix object from a .bed file, BEDMatrix-package to learn more about the BEDMatrix package, LinkedMatrix to link several BEDMatrix objects together.

Examples

```
# Get the path to the example .bed file
path <- system.file("extdata", "example.bed",
                    package = "BEDMatrix")

# Create a BEDMatrix object the example .bed file
m <- BEDMatrix(path)

# Get the dimensions of the BEDMatrix object
dim(m)

# Get the row names of the BEDMatrix object
rownames(m)

# Get the column names of the BEDMatrix object
colnames(m)

# Extract genotypes for the specified sample(s)
m[1, ]
m[1:3, ]
m["per0_per0", ]
m[c("per0_per0", "per1_per1", "per2_per2"), ]

# Extract genotypes for a particular variant
m[, 1]
m[, c("snp0_A", "snp1_C", "snp2_G")]

# Extract genotypes for the specified samples and variants
m[
    c("per0_per0", "per1_per1", "per2_per2"),
    c("snp0_A", "snp1_C", "snp2_G")
]
```

Index