

Package ‘BEDMatrix’

April 25, 2019

Version 1.5.0

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>

SystemRequirements C++11

Depends R (>= 3.0.0)

Imports methods, Rcpp (>= 0.12.1), crochet (>= 2.0.0)

LinkingTo Rcpp, BH

Suggests data.table, testthat, covr

RoxygenNote 6.1.1

Encoding UTF-8

NeedsCompilation yes

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

Maintainer Alexander Grueneberg <alexander.grueneberg@gmail.com>

Repository CRAN

Date/Publication 2019-04-25 15:00:03 UTC

R topics documented:

BEDMatrix-package	2
BEDMatrix-class	2
initialize,BEDMatrix-method	4
show,BEDMatrix-method	5

BEDMatrix-package	<i>A Package to Extract Genotypes from a PLINK .bed File.</i>
-------------------	---

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](#) to learn more about the `BEDMatrix` class.

BEDMatrix-class	<i>A Class to Extract Genotypes from a PLINK .bed File.</i>
-----------------	---

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 similarly to `.raw` files (generated with the `--recodeA` argument in `PLINK`): 0 indicates homozygous major allele, 1 indicates heterozygous, and 2 indicates homozygous minor allele.

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 the `Rcpp` package and the `Boost.Interprocess` library provided by the `BH` package.

Slots

- xptr An external pointer to the underlying [Rcpp](#) code.
- dims An integer vector specifying the number of samples and variants as determined by the accompanying [.fam](#) and [.bim](#) files or by the n and p parameters of the [constructor function](#).
- dimnames 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 [constructor function](#) were provided.
- path A character string containing the path to the [.bed](#) file.

Methods

- [
- dim
- dimnames
- dimnames<-
- as.matrix
- is.matrix
- length
- str
- show
- initialize

See Also

[initialize\(\)](#) to create a [BEDMatrix](#) object from a [.bed](#) file, [BEDMatrix-package](#) to learn more about [.bed](#) files, [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")
]

```

```
initialize,BEDMatrix-method
```

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

```

## S4 method for signature 'BEDMatrix'
initialize(.Object, path, n = NULL, p = NULL,
  simple_names = FALSE)

```

Arguments

.Object	Internal, used by <code>methods::initialize()</code> generic.
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, <code>_</code> , and within-family IDs, while column names are concatenations of variant names, <code>_</code> , and minor alleles. 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 **.bed** files.

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)

# Alternatively, a BEDMatrix object can also be created using the `new`
# function
m3 <- new("BEDMatrix", path = path)
```

show,BEDMatrix-method *Show a BEDMatrix Object.*

Description

Display the object, by printing, plotting or whatever suits its class.

Usage

```
## S4 method for signature 'BEDMatrix'
show(object)
```

Arguments

object A [BEDMatrix](#) object.

Index

BEDMatrix, [2, 4-6](#)
BEDMatrix (BEDMatrix-class), [2](#)
BEDMatrix-class, [2](#)
BEDMatrix-package, [2, 3, 5](#)
BH, [2](#)

constructor function, [3](#)

initialize(), [3](#)
initialize, BEDMatrix-method, [4](#)

LinkedMatrix, [3](#)

methods::initialize(), [4](#)

Rcpp, [2, 3](#)

show, BEDMatrix-method, [5](#)