Package 'BIGDAWG'

September 6, 2017

Type Package

Title Case-Control Analysis of Multi-Allelic Loci

Version 1.16

Date 2017-09-05

Author Derek Pappas <dpappas@chori.org>, Steve Mack <sjmack@chori.org>, Jill Hollenbach <Jill.Hollenbach@ucsf.edu>

Maintainer Steve Mack <sjmack@chori.org>

URL http://tools.immunogenomics.org/, https://github.com/IgDAWG/BIGDAWG

BugReports https://github.com/IgDAWG/BIGDAWG/issues

Description Data sets and functions for chi-squared Hardy-Weinberg and case-control association tests of highly polymorphic genetic data [e.g., human leukocyte antigen (HLA) data]. Performs association tests at multiple levels of polymorphism (haplotype, locus and HLA amino-acids) as described in Pappas DJ, Marin W, Hollenbach JA, Mack SJ (2016) <doi:10.1016/j.humimm.2015.12.006>. Combines rare variants to a common class to account for sparse cells in tables as described by Hollenbach JA, Mack SJ, Thomson G, Gourraud PA (2012) <doi:10.1007/978-1-61779-842-9_14>.

License GPL (>= 3)

LazyData True

Depends R (>= 3.3.2)

Imports XML, haplo.stats, parallel, knitr, rmarkdown

VignetteBuilder knitr

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-09-06 06:09:33 UTC

R topics documented:

`A` *Amino Acid Analysis Function*

Description

This is the workhorse function for the amino acid analysis.

Usage

```
A(loci.ColNames, Locus, genos, grp, nGrp0, nGrp1, ExonAlign, Cores)
```

Arguments

`loci.ColNames`	The column names of the loci being analyzed.
`Locus`	Locus being analyzed.
`genos`	Genotype table.
`grp`	Case/Control or Phenotype groupings.
`nGrp0`	Number of controls.
`nGrp1`	Number of cases.
`ExonAlign`	Exon protein alignment filtered for locus.
`Cores`	Number of cores to use for analysis.

Note

This function is for internal BIGDAWG use only.

A.wrapper *Amino Acid Wrapper*

Description

Wrapper function for amino acid analysis.

Usage

```
A.wrapper(nloci, loci, loci.ColNames, genos, grp, nGrp0, nGrp1, EPL, Cores,
  Output, Verbose)
```

Arguments

nloci	Number of loci being analyzed.
loci	Loci being analyzed.
loci.ColNames	The column names of the loci being analyzed.
genos	Genotype table.
grp	Case/Control or Phenotype groupings.
nGrp0	Number of controls.
nGrp1	Number of cases.
EPL	Exon protein alignment.
Cores	Number of cores to use for analysis.
Output	Data return carryover from main BIGDAWG function
Verbose	Summary display carryover from main BIGDAWG function

Note

This function is for internal BIGDAWG use only.

AA.df.check *Contingency Table Check*

Description

Checks amino acid contingency table data frame to ensure required variation exists.

Usage

```
AA.df.check(x)
```

Arguments

x	contingency table.

Note

This function is for internal BIGDAWG use only.

AAtable.builder *Amino Acid Contingency Table Build*

Description

Build Contingency Tables for Amino Acid Analysis.

Usage

```
AAtable.builder(x, y)
```

Arguments

x	Filtered alignmnet list element.
y	Phenotype groupings.

Note

This function is for internal BIGDAWG use only.

AlignmentFilter *Alignment Filter*

Description

Filter Protein Exon Alignment File for Specific Alleles.

Usage

```
AlignmentFilter(Align, Alleles, Locus)
```

Arguments

Align	Protein Alignment Object.
Alleles	to be pulled.
Locus	Locus to be filtered against.

Note

This function is for internal BIGDAWG use only.

AlignObj.Create *Alignment Object Creator*

Description

Synthesize Object for Exon Protein Alignments.

Usage

```
AlignObj.Create(Loci, Release, RefTab)
```

Arguments

Loci	Loci to be bundled.
Release	IMGT/HLA database release version.
RefTab	Data of reference exons used for protein alignment creation.

Note

This function is for internal BIGDAWG use only.

AlignObj.Update *Updated Alignment Object Creator*

Description

Synthesize Object for Exon Protein Alignments.

Usage

```
AlignObj.Update(Loci, Release, RefTab)
```

Arguments

Loci	Loci to be bundled.
Release	IMGT/HLA database release version.
RefTab	Data of reference exons used for protein alignment creation.

Note

This function is for internal BIGDAWG use only.

BIGDAWG *BIGDAWG Main Wrapper Function*

Description

This is the main wrapper function for each analysis.

Usage

```
BIGDAWG(Data, HLA = TRUE, Run.Tests, Loci.Set, All.Pairwise = FALSE,
  Trim = FALSE, Res = 2, EVS.rm = FALSE, Missing = 2, Cores.Lim = 1L,
  Results.Dir, Return = FALSE, Output = TRUE, Merge.Output = FALSE,
  Verbose = TRUE)
```

Arguments

Data	Name of the genotype data file.
HLA	Logical indicating whether data is HLA class I/II genotyping data only.
Run.Tests	Specifics which tests to run.
Loci.Set	Input list defining which loci to use for analyses (combinations permitted).
All.Pairwise	Logical indicating whether all pairwise loci should be analyzed in haplotype analysis.
Trim	Logical indicating if HLA alleles should be trimmed to a set resolution.
Res	Numeric setting what desired resolution to trim HLA alleles.
EVS.rm	Logical indicating if expression variant suffixes should be removed.
Missing	Numeric setting allowable missing data for running analysis (may use "ignore").
Cores.Lim	Interger setting the number of cores accessible to BIGDAWG (Windows limit is 1 core).
Results.Dir	Optional, string of full path directory name for BIGDAWG output.
Return	Logical Should analysis results be returned as list.
Output	Logical Should analysis results be written to output directory.
Merge.Output	Logical Should analysis results be merged into a single file for easy access.
Verbose	Logical Should a summary of each analysis be displayed in console.

Examples

```
### The following examples use the synthetic data set bundled with BIGDAWG

# Haplotype analysis with no missing genotypes for two loci sets
# Significant haplotypic association with phenotype
# BIGDAWG(Data="HLA_data", Run.Tests="H", Missing=0, Loci.Set=list(c("DRB1","DQB1")))

# Hardy-Weinberg and Locus analysis ignoring missing data
```

```
# Significant locus associations with phenotype at all but DQB1
# BIGDAWG(Data="HLA_data", Run.Tests="L", Missing="ignore")

# Hardy-Weinberg analysis trimming data to 2-Field resolution
# Significant locus deviation at DQB1
BIGDAWG(Data="HLA_data", Run.Tests="HWE", Trim=TRUE, Res=2)
```

`buildHAPnames` *Haplotype Name Builder*

Description

Builds table of names for HAPsets

Usage

```
buildHAPnames(Combn, loci)
```

Arguments

`Combn`	Combination of loci to extraction from genos
`loci`	Character vector of unique loci being analyzed.

Note

This function is for internal BIGDAWG use only.

`buildHAPsets` *Haplotype List Builder*

Description

Builds table of haplotypes from combinations

Usage

```
buildHAPsets(Combn, genos, loci, loci.ColNames)
```

Arguments

`Combn`	Combination of loci to extraction from genos
`genos`	The genotype columns of the loci set being analyzed.
`loci`	Character vector of unique loci being analyzed.
`loci.ColNames`	Character vector of genos column names.

Note

This function is for internal BIGDAWG use only.

cci *Case-Control Odds ratio calculation and graphing*

Description

cci function port epicalc version 2.15.1.0 (Virasakdi Chongsuvivatwong, 2012)

Usage

```
cci(caseexp, controlex, casenonex, controlnonex, cctable = NULL,
  graph = TRUE, design = "cohort", main, xlab, ylab, xaxis, yaxis,
  alpha = 0.05, fisher.or = FALSE, exact.ci.or = TRUE, decimal = 2)
```

Arguments

caseexp	Number of cases exposed
controlex	Number of controls exposed
casenonex	Number of cases not exosed
controlnonex	Number of controls not exposed
cctable	A 2-by-2 table. If specified, will supercede the outcome and exposure variables
graph	If TRUE (default), produces an odds ratio plot
design	Specification for graph; can be "case control","case-control", "cohort" or "prospective"
main	main title of the graph
xlab	label on X axis
ylab	label on Y axis
xaxis	two categories of exposure in graph
yaxis	two categories of outcome in graph
alpha	level of significance
fisher.or	whether odds ratio should be computed by the exact method
exact.ci.or	whether confidence limite of the odds ratio should be computed by the exact method
decimal	number of decimal places displayed

Note

This function is for internal BIGDAWG use only.

cci.pval *Case Control Odds Ratio Calculation from Epicalc*

Description

Calculates odds ratio and pvalues from 2x2 table

Usage

```
cci.pval(x)
```

Arguments

x	List of 2x2 matrices for calculation, output of TableMaker.

Note

This function is for internal BIGDAWG use only.

cci.pval.list *Case Control Odds Ratio Calculation from Epicalc list variation*

Description

Variation of the cci.pvalue function

Usage

```
cci.pval.list(x)
```

Arguments

x	List of 2x2 matrices to apply the cci.pvalue function. List output of TableMaker.

Note

This function is for internal BIGDAWG use only.

Check.Cores *Check Cores Parameters*

Description

Check cores limitation for OS compatibility

Usage

```
Check.Cores(Cores.Lim)
```

Arguments

Cores.Lim	Integer How many cores can be used.

Check.Params *Check Input Parameters*

Description

Check input parameters for invalid entries.

Usage

```
Check.Params(HLA, All.Pairwise, Trim, Res, EVS.rm, Missing, Cores.Lim, Return,
  Output, Merge.Output, Verbose)
```

Arguments

HLA	Logical indicating whether data is HLA class I/II genotyping data only.
All.Pairwise	Logical indicating whether all pairwise loci should be analyzed in haplotype analysis.
Trim	Logical indicating if HLA alleles should be trimmed to a set resolution.
Res	Numeric setting what desired resolution to trim HLA alleles.
EVS.rm	Logical indicating if expression variant suffixes should be removed.
Missing	Numeric setting allowable missing data for running analysis (may use "ignore").
Cores.Lim	Interger setting the number of cores accessible to BIGDAWG (Windows limit is 1 core).
Return	Logical Should analysis results be returned as list.
Output	Logical Should analysis results be written to output directory.
Merge.Output	Logical Should analysis results be merged into a single file for easy access.
Verbose	Logical Should a summary of each analysis be displayed in console.

Note

This function is for internal use only.

CheckAlleles *Allele Legitimacy Check*

Description

Checks available alleles against data to ensure complete overlap.

Usage

```
CheckAlleles(x, y, z1, z2)
```

Arguments

x	Exon protein list alignment object.
y	Genotypes from data file
z1	loci in data file
z2	Genotype column names

Note

This function is for internal BIGDAWG use only.

CheckHLA *HLA Formatting Check*

Description

Checks data to see if HLA data is properly formatted.

Usage

```
CheckHLA(x)
```

Arguments

x	All columns of HLA genotyping data.

Note

This function is for internal BIGDAWG use only.

CheckLoci *Loci Legitimacy Check*

Description

Checks available loci against data to ensure complete overlap.

Usage

```
CheckLoci(x, y)
```

Arguments

x	Loci available in exon protein list alignment object.
y	Unique column names

Note

This function is for internal BIGDAWG use only.

CheckRelease *Function to Check Release Versions*

Description

This updates the protein aligment used in checking HLA loci and alleles as well as in the amino acid analysis.

Usage

```
CheckRelease(Package = T, Alignment = T, Output = F)
```

Arguments

Package	Logical to check for BIGDAWG package versions
Alignment	Logical to check the IMGT/HLA database version for the alignment bundled with BIGDAWG.
Output	Should any error be written to a file

Note

Requires active internet connection.

DRB345.flag *DRB345 haplotype zygosity flag check*

Description

Identify DR345 flagged haplotypes

Usage

```
DRB345.flag(tmp, Tab)
```

Arguments

tmp	Output of DRB345.zygosity
Tab	Data frame of sampleIDs, phenotypes, and genotypes

Note

This function is for internal BIGDAWG use only.

DRB345.parser *DRB345 Column Processing*

Description

Separates DRB345 column pair into separate columns for each locus

Usage

```
DRB345.parser(Tab)
```

Arguments

Tab	Data frame of sampleIDs, phenotypes, and genotypes

Note

This function is for internal BIGDAWG use only.

DRB345.zygosity *DRB345 haplotype zygosity checker*

Description

Checks DR haplotypes for correct zygosity and flags unanticipated haplotypes

Usage

```
DRB345.zygosity(x)
```

Arguments

x	Row of data set data frame following DRB345 parsing

Note

This function is for internal BIGDAWG use only.

Err.Log *Error Code Display and Logging*

Description

Displays error codes attributable to data formatting and Locus/Allele naming. Writes to log file.

Usage

```
Err.Log(Output, x, y = NULL)
```

Arguments

Output	Logical indicating if Error logging should be written to a file.
x	Log Code.
y	Misc information relevant to error.

Note

This function is for internal BIGDAWG use only.

EVSremoval *Expression Variant Suffix Removal*

Description

Removes expression variant suffixes from HLA alleles in the exon protein alignment object.

Usage

```
EVSremoval(Locus, EPList)
```

Arguments

Locus	Locus to be filtered against.
EPList	Exon Protein Alignment Object

Note

This function is for internal BIGDAWG use only.

ExonPtnAlign.Create *Protein Exon Alignment Formatter*

Description

Dynamically creates an alignmnet of Allele exons for Analysis.

Usage

```
ExonPtnAlign.Create(Locus, RefTab)
```

Arguments

Locus	Locus alignment to be formatted.
RefTab	Reference exon protein information for alignment formatting.

Note

This function is for internal BIGDAWG use only.

ExonPtnList *Exon 2 (class II) or 2/3 (class I) protein alignments.*

Description

Alignment object for use in the amino acid analysis.

Usage

```
ExonPtnList
```

Format

A list wherein each element is an alignment dataframe for a single locus.

FixAlleleName *Allele Name Format Fix*

Description

Separate locus and allele names if data is formatted as Loci*Allele

Usage

```
FixAlleleName(Output, Tab)
```

Arguments

Output	Logical indicating if Error logging should be written to a file.
Tab	All columns of HLA genotyping data.

Note

This function is for internal BIGDAWG use only.

getAllele.Count *Recompute number of alleles*

Description

Using Freq.Final, recompute number of alleles

Usage

```
getAllele.Count(x)
```

Arguments

x	Locus specific contingency matrix getCS.Mat output.

Note

This function is for internal BIGDAWG use only.

getCS.Mat *Chi square matrices*

Description

Chi Square contingency matrix builder with rare cell binning

Usage

```
getCS.Mat(Locus, genos.sub, Allele.Freq, Allele.Combn)
```

Arguments

Locus	Locus of interest.
genos.sub	Genotypes for locus of interest.
Allele.Freq	Allele frequencies.
Allele.Combn	Allele combinations.

Note

This function is for internal BIGDAWG use only.

getCS.stat *Chi square test statistic*

Description

Calculate chi square test statistic

Usage

```
getCS.stat(Locus, Freq.Final)
```

Arguments

Locus	Locus of interest.
Freq.Final	Contingency Matrix getCS.Mat output.

Note

This function is for internal BIGDAWG use only.

GetField *HLA trimming function*

Description

Trim a properly formatted HLA allele to desired number of fields.

Usage

```
GetField(x, Res)
```

Arguments

x	HLA allele.
Res	Resolution desired.

Note

This function is for internal BIGDAWG use only.

GetFiles *File Fetcher*

Description

Download Protein Alignment and Accessory Files

Usage

```
GetFiles(Loci)
```

Arguments

Loci	HLA Loci to be fetched. Limited Loci available.

Note

This function is for internal BIGDAWG use only.

getHap *Haplotype Table Maker*

Description

Builds table of haplotypes

Usage

```
getHap(SID, HaploEM)
```

Arguments

SID	Index number (i.e., row number) of sample ID from genotype matrix.
HaploEM	Haplotype output object from haplo.stat::haplo.em function.

Note

This function is for internal BIGDAWG use only.

getObsFreq *Observed Frequency*

Description

Get observed frequency of genotypes

Usage

```
getObsFreq(x, genos.locus)
```

Arguments

x	Single genotype.
genos.locus	Locus genotypes.

Note

This function is for internal BIGDAWG use only.

H.MC *Haplotype Analysis Function for Multicore*

Description

This is the workhorse function for the haplotype analysis.

Usage

```
H.MC(genos.sub, grp, Verbose)
```

Arguments

genos.sub	The genotype columns of the loci(locus) set being analyzed.
grp	Case/Control or Phenotype groupings.
Verbose	Summary display carryover from main BIGDAWG function

Note

This function is for internal BIGDAWG use only.

H.MC.wrapper *Haplotype Wrapper for Multicore*

Description

Wrapper for main H function

Usage

```
H.MC.wrapper(SID, Tabsub, loci, loci.ColNames, genos, grp, All.Pairwise, Output,
  Verbose, Cores)
```

Arguments

SID	Character vector of subject IDs.
Tabsub	Data frame of genotype calls for set being analyzed.
loci	Character vector of unique loci being analyzed.
loci.ColNames	Character vector of genos column names.
genos	The genotype columns of the loci set being analyzed.
grp	Case/Control or Phenotype groupings.
All.Pairwise	Haplotype argument carryover from main BIGDAWG function
Output	Data return carryover from main BIGDAWG function
Verbose	Summary display carryover from main BIGDAWG function
Cores	Cores carryover from main BIGDAWG function

Note

This function is for internal BIGDAWG use only.

HLA_data *Example HLA Dataset*

Description

A synthetic dataset of HLA genotypes for using bigdawg.

Usage

```
HLA_data
```

Format

A data frame with 2000 rows and 14 variables

HWE *Hardy Weinbergy Equilibrium Function*

Description

This is the main function for the HWE analysis.

Usage

```
HWE(Tab, All.ColNames)
```

Arguments

Tab	data frame of genotype files post processing.
All.ColNames	character vector of Tab object column names.

Note

This function is for internal BIGDAWG use only.

HWE.ChiSq *Hardy Weinbergy Equilibrium Function*

Description

This is the workhorse function for each group analysis.

Usage

```
HWE.ChiSq(genos.sub, loci, nloci)
```

Arguments

genos.sub	data frame of genotype files post processing.
loci	list of loci.
nloci	number of loci in list

Note

This function is for internal BIGDAWG use only.

HWE.wrapper *Hardy-Weinbery Wrapper*

Description

Wrapper for main HWE function

Usage

```
HWE.wrapper(Tab, All.ColNames, Output, Verbose)
```

Arguments

Tab	Data frame of genotype files post processing.
All.ColNames	Character vector of Tab object column names.
Output	Data return carryover from main BIGDAWG function
Verbose	Summary display carryover from main BIGDAWG function

Note

This function is for internal BIGDAWG use only.

L *Locus Analysis Function*

Description

This is the workhorse function for the locus level analysis.

Usage

```
L(loci.ColNames, Locus, genos, grp, nGrp0, nGrp1)
```

Arguments

loci.ColNames	The column names of the loci being analyzed.
Locus	Locus being analyzed.
genos	Genotype table
grp	Case/Control or Phenotype groupings.
nGrp0	Number of controls
nGrp1	Number of cases

Note

This function is for internal BIGDAWG use only.

L.wrapper *Locus Wrapper*

Description

Wrapper for main L function

Usage

```
L.wrapper(nloci, loci, loci.ColNames, genos, grp, nGrp0, nGrp1, Output, Verbose)
```

Arguments

nloci	Number of loci being analyzed.
loci	Loci being analyzed.
loci.ColNames	The column names of the loci being analyzed.
genos	Genotype table
grp	Case/Control or Phenotype groupings.
nGrp0	Number of controls
nGrp1	Number of cases
Output	Data return carryover from main BIGDAWG function
Verbose	Summary display carryover from main BIGDAWG function

Note

This function is for internal BIGDAWG use only.

make2x2 *Creation of a 2x2 table using the indicated orientation.*

Description

make2x2 function port epicalc version 2.15.1.0 (Virasakdi Chongsuvivatwong, 2012)

Usage

```
make2x2(caseexp, controlex, casenonex, controlnonex)
```

Arguments

caseexp	Number of cases exposed
controlex	Number of controls exposed
casenonex	Number of cases not exosed
controlnonex	Number of controls not exposed

Note

This function is for internal BIGDAWG use only.

---

makeComb *Genotype Combination Maker*

Description

Make data frame of possible genotype combinations

Usage

```
makeComb(x)
```

Arguments

x	Number of alleles.

Note

This function is for internal BIGDAWG use only.

MergeData_Output *Data Object Merge and Output*

Description

Whole data set table construction of per haplotype for odds ratio, confidence intervals, and pvalues

Usage

```
MergeData_Output(BD.out, Run, OutDir)
```

Arguments

BD.out	Output of analysis as list.
Run	Tests that are to be run as defined by Run.Tests.
OutDir	Output directory defined by Results.Dir or default.

Note

This function is for internal BIGDAWG use only.

PgrpExtract *HLA P group Finder*

Description

Identify P group for a given allele if exists.

Usage

```
PgrpExtract(x, y)
```

Arguments

x	Allele of interest.
y	Formatted P groups.

Note

This function is for internal BIGDAWG use only.

PgrpFormat *HLA P group File Formatter*

Description

Format the hla_nom_p.txt read table object for a specific locus.

Usage

```
PgrpFormat(x, Locus)
```

Arguments

x	P group object from read.table command.
Locus	Locus to be filtered on.

Note

This function is for internal BIGDAWG use only.

PreCheck *Data Summary Function*

Description

Summary function for sample population within data file.

Usage

```
PreCheck(Tab, All.ColNames, rescall, HLA, Verbose, Output)
```

Arguments

Tab	Loci available in exon protein list alignment object.
All.ColNames	Column names from genotype data.
rescall	HLA resolution set for analysis.
HLA	HLA bigdawg argument passed to function
Verbose	Summary display carryover from BIGDAWG function.
Output	Data output carryover form BIGDAWG function

Note

This function is for internal BIGDAWG use only.

rmABstrings *Replace absent allele strings*

Description

Replaces allowable absent allele strings with ^ symbol.

Usage

```
rmABstrings(df)
```

Arguments

df	Genotypes dataframe.

Note

This function is for internal BIGDAWG use only.

RunChiSq *Chi-squared Contingency Table Test*

Description

Calculates chi-squared contingency table tests and bins rare cells.

Usage

```
RunChiSq(x)
```

Arguments

x	Contingency table.

Note

This function is for internal BIGDAWG use only.

summaryGeno.2 *Haplotype missing Allele summary function*

Description

Summary function for identifying missing alleles in a matrix of genotypes.

Usage

```
summaryGeno.2(geno, miss.val = 0)
```

Arguments

geno	Matrix of genotypes.
miss.val	Vector of codes for allele missing values.

Note

This function is for internal BIGDAWG use only and is ported from haplo.stats.

TableMaker *Table Maker*

Description

Table construction of per haplotype for odds ratio, confidence intervals, and pvalues

Usage

```
TableMaker(x)
```

Arguments

x	Contingency table with binned rare cells.

Note

This function is for internal BIGDAWG use only.

UpdateRelease *Update function for protein aligment upon new IMGT HLA data release*

Description

This updates the protein aligment used in checking HLA loci and alleles as well as in the amino acid analysis. Alignment must exist in database (ftp://ftp.ebi.ac.uk/pub/databases/ipd/imgt/hla/alignments/) or update will fail.

Usage

```
UpdateRelease(Add.Loci = NULL, Force = F, Restore = F, Output = F)
```

Arguments

Add.Loci	Character string or vector of loci that should be added to default loci (default = HLA-A,B,C,DRB1/3/4/5,DQA1,DQB1,DPA1,DPB1).
Force	Logical specifiying if update should be forced.
Restore	Logical specifying if the original alignment file be restored.
Output	Logical indicating if error reporting should be written to file.

Index