Package 'dartR.sexlinked'

Type Package

Title Analysing SNP Data to Identify Sex-Linked Markers

Version 1.0.5

Date 2024-06-22

Author Diana Robledo-Ruiz [aut, cre],
Floriaan Devloo-Delva [aut],
Bernd Gruber [aut],
Arthur Georges [aut],
Jose L. Mijangos [aut],
Carlo Pacioni [aut],
Peter J. Unmack [ctb],
Oliver Berry [ctb]

Description Identifies, filters and exports sex linked markers using 'SNP'
(single nucleotide polymorphism) data. To install the other packages, we
recommend to install the 'dartRverse' package, that supports the installation of all pack-
ages in the 'dartRverse'.
If you want understand the applied rational to identify sexlinked markers
and/or want to cite 'dartR.sexlinked', you find the information by typing
citation('dartR.sexlinked') in the console.

License GPL (>= 3)

URL https://green-striped-gecko.github.io/dartR/

BugReports https://groups.google.com/g/dartr?pli=1

Depends adegenet (>= 2.0.0), dartR.base, dartR.data, R (>= 3.5)

Imports doParallel (>= 1.0.17), ggplot2, methods, utils, foreach,
patchwork

Encoding UTF-8

NeedsCompilation no

RoxygenNote 7.3.1

Maintainer Diana Robledo-Ruiz <diana.robledoruiz1@monash.edu>

Repository CRAN

Date/Publication 2024-06-24 15:40:02 UTC

Contents

---

gl.drop.sexlinked   *Removes loci that are sex linked*

Description

This function identifies sex-linked and autosomal loci present in a SNP dataset (genlight object) using individuals with known sex. It identifies five types of loci: w-linked or y-linked, sex-biased, z-linked or x-linked, gametologous and autosomal.

This function produces as output a genlight object with autosomal loci only.

Usage

```
gl.drop.sexlinked(
  x,
  system = NULL,
  ncores = 1,
  plot.display = TRUE,
  plot.theme = theme_dartR(),
  plot.colors = NULL,
  plot.file = NULL,
  plot.dir = NULL,
  verbose = NULL
)
```

Arguments

x	Name of the genlight object containing the SNP data. This genlight object needs to contain the sex of the individuals. See explanation in details [required].
system	String that declares the sex-determination system of the species: 'zw' or 'xy' [required].
ncores	Number of processes to be used in parallel operation. If ncores > 1 parallel operation is activated, see "Details" section [default 1].
plot.display	Creates four output plots. See explanation in details [default TRUE].
plot.theme	Theme for the plot. See Details for options [default theme_dartR()].[not yet implemented]
plot.colors	[not implemented yet]

plot.file	Name for the RDS binary file to save (base name only, exclude extension) [default NULL].
plot.dir	Directory to save the plot RDS files [default as specified by the global working directory or tempdir()].
verbose	Verbosity: 0, silent or fatal errors; 1, begin and end; 2, progress log; 3, progress and results summary; 5, full report [default NULL, unless specified using gl.set.verbosity].

Details

The genlight object must contain in `gl@other$ind.metrics` a column named "id", and a column named "sex" in which individuals with known-sex are assigned 'M' for male, or 'F' for female. The function ignores individuals that are assigned anything else or nothing at all (unknown-sex).

The creation of plots can be turned-off (`plot.display = FALSE`) in order to save a little bit of running time for very large datasets (>50,000 SNPs). However, we strongly encourage you to always inspect the output plots at least once to make sure everything is working properly.

Function's output

This function returns as output a genlight object that contains only autosomal loci (i.e. sex-linked loci have been dropped.)

And four plots:

- A BEFORE plot based on loci call rate by sex, with w/y-linked loci colored in yellow and sex-biased loci in blue
- An AFTER plot based on loci call rate by sex, with sex-linked loci removed
- A BEFORE plot based on loci heterozygosity by sex, with z/x-linked loci colored in orange and gametologs in green
- An AFTER plot based on loci heterozygosity by sex, with sex-linked loci removed

Value

A genlight object and 4 plots.

Author(s)

Custodian: Diana Robledo-Ruiz – Post to https://groups.google.com/d/forum/dartr

References

- Robledo-Ruiz, D. A., Austin, L., Amos, J. N., Castrejón-Figueroa, J., Harley, D. K., Magrath, M. J., Sunnucks, P., & Pavlova, A. (2023). Easy-to-use R functions to separate reduced-representation genomic datasets into sex-linked and autosomal loci, and conduct sex assignment. Molecular Ecology Resources, 00, 1-21.

Examples

```
LBP_noSexLinked <- gl.drop.sexlinked(x = LBP, system = "xy", plot.display = TRUE, ncores = 1)
LBP_noSexLinked
```

---

gl.infer.sex   *Uses sex-linked loci to infer sex of individuals*

Description

This function uses the output of function gl.keep.sexlinked (list of 5 elements) to infer the sex of all individuals in the dataset. It uses 3 types of sex-linked loci (W-/Y-linked, Z-/X-linked, and gametologs), assigns a preliminary genetic sex for each type of sex-linked loci available, and outputs an agreed sex.

This function produces as output a dataframe with individuals in rows and 11 columns.

Usage

```
gl.infer.sex(gl_sexlinked, system = NULL, seed = NULL)
```

Arguments

gl_sexlinked	The output of function gl.keep.sexlinked (complete list with 5 elements). See explanation in "Details" section [required].
system	String that declares the sex-determination system of the species: 'zw' or 'xy' [required].
seed	User-defined integer for repeatability purposes. If not provided by user, it is chosen randomly by default. See "Details" section.

Details

Parameter `gl_sexlinked` must be the name of the output object (a list of 5 elements) produced by function `gl.keep.sexlinked`. Parameter seed must be an integer that will be used on the KMeans algorithm used by the function. We highly recommend choosing the seed to guarantee repeatability.

Note that this function was created with the explicit intent that a human checks the evidence for the sex assignments that do NOT agree for all types of sex-linked loci (called "indefinite sex assignments" and denoted as "*M" or "*F" in the last column of dataframe output). This human can then use their criterion to validate these assignments.

Function's output

This function creates a dataframe with one row per individual and 11 columns:

- id > Individuals' ID.
- w.linked.sex or y.linked.sex > Sex inferred using w-linked or y-linked loci.
- #called > Number of W-linked or Y-linked loci for which the individual had a called genotype (cf. missing genotype).
- #missing > Number of W-linked or Y-linked loci for which the individual had a missing genotype (cf. called genotype).
- z.linked.sex or x.linked.sex > Sex inferred using z-linked or x-linked loci.

- #Hom.z or #Hom.x > Number of z-linked or x-linked loci for which the individual is homozygous.

- #Het.z or #Het.x > Number of z-linked or x-linked loci for which the individual is heterozygous.

- gametolog.sex > Sex inferred using gametologs.

- #Hom.g > Number of gametologous loci for which the individual is homozygous.

- #Het.g > Number of gametologous loci for which the individual is heterozygous.

- agreed.sex > Agreed sex: 'F' or 'M' if all preliminary sex-assignments match (i.e., definite sex assignment), and '*F' or '*M' if NOT all preliminary sex-assignments match (i.e., indefinite sex assignment).

Value

A dataframe.

Author(s)

Custodian: Diana Robledo-Ruiz – Post to <https://groups.google.com/d/forum/dartr>

References

- Robledo-Ruiz, D. A., Austin, L., Amos, J. N., Castrejón-Figueroa, J., Harley, D. K., Magrath, M. J., Sunnucks, P., & Pavlova, A. (2023). Easy-to-use R functions to separate reduced-representation genomic datasets into sex-linked and autosomal loci, and conduct sex assignment. Molecular Ecology Resources, 00, 1-21.

Examples

```
LBP_sexLinked <- gl.keep.sexlinked(x = LBP, system = "xy", plot.display = TRUE, ncores = 1)
inferred.sexes <- gl.infer.sex(gl_sexlinked = LBP_sexLinked, system = "xy", seed = 100)
inferred.sexes
```

---

gl.keep.sexlinked *Keeps loci that are sex linked*

Description

This function identifies sex-linked and autosomal loci present in a SNP dataset (genlight object) using individuals with known sex. It identifies five types of loci: w-linked or y-linked, sex-biased, z-linked or x-linked, gametologous and autosomal.

This function produces as output a list with 5 elements, including one dataframe and 4 genlight objects with sex-linked loci.

Usage

```
gl.keep.sexlinked(
  x,
  system = NULL,
  ncores = 1,
  plot.display = TRUE,
  plot.theme = theme_dartR(),
  plot.colors = NULL,
  plot.file = NULL,
  plot.dir = NULL,
  verbose = NULL
)
```

Arguments

x	Name of the genlight object containing the SNP data. This genlight object needs to contain the sex of the individuals. See explanation in details [required].
system	String that declares the sex-determination system of the species: 'zw' or 'xy' [required].
ncores	Number of processes to be used in parallel operation. If ncores > 1 parallel operation is activated, see "Details" section [default 1].
plot.display	Creates four output plots. See explanation in details [default TRUE].
plot.theme	Theme for the plot. See Details for options [default theme_dartR()].[not yet implemented]
plot.colors	[not implemented yet]
plot.file	Name for the RDS binary file to save (base name only, exclude extension) [default NULL].
plot.dir	Directory to save the plot RDS files [default as specified by the global working directory or tempdir()].
verbose	Verbosity: 0, silent or fatal errors; 1, begin and end; 2, progress log; 3, progress and results summary; 5, full report [default NULL, unless specified using gl.set.verbosity].

Details

The genlight object must contain in `gl@other$ind.metrics` a column named "id", and a column named "sex" in which individuals with known-sex are assigned 'M' for male, or 'F' for female. The function ignores individuals that are assigned anything else or nothing at all (unknown-sex).

The creation of plots can be turned-off (`plot.display = FALSE`) in order to save a little bit of running time for very large datasets (>50,000 SNPs). However, we strongly encourage you to always inspect the output plots at least once to make sure everything is working properly.

Function's output

This function returns a list of 5 elements:

- $results.table > Table with statistics (columns) for each loci (rows)
- $w.linked or $y.linked > Genlight object with w-linked/y-linked loci

- $sex.biased > Genlight object with sex-biased scoring rate loci
- $z.linked or $x.linked > Genlight object with z-linked/x-linked loci
- $gametolog > Genlight object with gametologs

And four plots:

- A BEFORE plot based on loci call rate by sex, with w/y-linked loci colored in yellow and sex-biased loci in blue
- An AFTER plot based on loci call rate by sex, with only sex-linked loci
- A BEFORE plot based on loci heterozygosity by sex, with z/x-linked loci colored in orange and gametologs in green
- An AFTER plot based on loci heterozygosity by sex, with only sex-linked loci

Value

A list of 5 elements and 4 plots.

Author(s)

Custodian: Diana Robledo-Ruiz – Post to https://groups.google.com/d/forum/dartr

References

- Robledo-Ruiz, D. A., Austin, L., Amos, J. N., Castrejón-Figueroa, J., Harley, D. K., Magrath, M. J., Sunnucks, P., & Pavlova, A. (2023). Easy-to-use R functions to separate reduced-representation genomic datasets into sex-linked and autosomal loci, and conduct sex assignment. Molecular Ecology Resources, 00, 1-21.

Examples

```
LBP_sexLinked <- gl.keep.sexlinked(x = LBP, system = "xy", plot.display = TRUE, ncores = 1)
LBP_sexLinked$gametolog
```

---

gl.report.sexlinked   *Filters loci that are sex linked*

Description

This function identifies sex-linked and autosomal loci present in a SNP dataset (genlight object) using individuals with known sex. It identifies five types of loci: w-linked or y-linked, sex-biased, z-linked or x-linked, gametologous and autosomal.

This function produces as output a dataframe and 2 plots.

Usage

```
gl.report.sexlinked(
  x,
  system = NULL,
  ncores = 1,
  plot.display = TRUE,
  plot.theme = theme_dartR(),
  plot.colors = NULL,
  plot.file = NULL,
  plot.dir = NULL,
  verbose = NULL
)
```

Arguments

x	Name of the genlight object containing the SNP data. This genlight object needs to contain the sex of the individuals. See explanation in details [required].
system	String that declares the sex-determination system of the species: 'zw' or 'xy' [required].
ncores	Number of processes to be used in parallel operation. If ncores > 1 parallel operation is activated, see "Details" section [default 1].
plot.display	Creates two output plots. See explanation in details [default TRUE].
plot.theme	Theme for the plot. See Details for options [default theme_dartR()].[not yet implemented]
plot.colors	[not implemented yet]
plot.file	Name for the RDS binary file to save (base name only, exclude extension) [default NULL].
plot.dir	Directory to save the plot RDS files [default as specified by the global working directory or tempdir()].
verbose	Verbosity: 0, silent or fatal errors; 1, begin and end; 2, progress log; 3, progress and results summary; 5, full report [default NULL, unless specified using gl.set.verbosity].

Details

The genlight object must contain in gl@other$ind.metrics a column named "id", and a column named "sex" in which individuals with known-sex are assigned 'M' for male, or 'F' for female. The function ignores individuals that are assigned anything else or nothing at all (unknown-sex).

The creation of plots can be turned-off (plot.display = FALSE) in order to save a little bit of running time for very large datasets (>50,000 SNPs). However, we strongly encourage you to always inspect the output plots at least once to make sure everything is working properly.

Function's output

This function returns two plots:

- A plot based on loci call rate by sex, with w/y-linked loci colored in yellow and sex-biased loci in blue

- A plot based on loci heterozygosity by sex, with z/x-linked loci colored in orange and game-tologs in green

And a dataframe in which loci are in rows, and columns:

- index - Index number to identify loci
- count.F.miss - Count of females that have this locus as missing data (NA).
- count.M.miss - Count of males that have this locus as missing data (NA)
- count.F.scored - Count of females that have this locus scored (0, 1 or 2; i.e. non-missing)
- count.M.scored - Count of males that have this locus scored (0, 1 or 2; i.e. non-missing)
- ratio - Fisher's exact test estimate testing for the independence of call rate and sex for this locus
- p.value - P-value for the Fisher's exact test estimate
- p.adjusted - P-value adjusted for false discovery rate
- scoringRate.F - Female call rate (proportion of females that were scored for this locus; x-axis in the 1st plot)
- scoringRate.M - Male call rate (proportion of males that were scored for this locus; y-axis in the 1st plot)
- w.linked/y.linked - Boolean for this locus being w-linked/y-linked
- sex.biased - Boolean for this locus having sex-biased call rate
- count.F.het - Count of females that are heterozygous for this locus
- count.M.het - Count of males that are heterozygous for this locus
- count.F.hom - Count of females that are homozygous for this locus
- count.M.hom - Count of males that are homozygous for this locus
- stat - Fisher's exact test estimate testing for the independence of heterozygosity and sex for this locus
- stat.p.value - P-value for the Fisher's exact test estimate
- stat.p.adjusted - P-value adjusted for false discovery rate
- heterozygosity.F - Proportion of females that are heterozygotes for this locus (x-axis in the 2nd plot)
- heterozygosity.M - Proportion of males that are heterozygotes for this locus (y-axis in the 2nd plot)
- z.linked/x.linked - Boolean for this locus being z-linked/x.linked
- gametolog - Boolean for this locus being a gametolog

Value

A dataframe and 2 plots.

Author(s)

Custodian: Diana Robledo-Ruiz – Post to https://groups.google.com/d/forum/dartr

References

- Robledo-Ruiz, D. A., Austin, L., Amos, J. N., Castrejón-Figueroa, J., Harley, D. K., Magrath, M. J., Sunnucks, P., & Pavlova, A. (2023). Easy-to-use R functions to separate reduced-representation genomic datasets into sex-linked and autosomal loci, and conduct sex assignment. Molecular Ecology Resources, 00, 1-21.

Examples

```
out <- gl.report.sexlinked(x = LBP, system = "xy", plot.display = TRUE, ncores = 1)
```

Index