Package 'hiphop'

August 19, 2020

Title Parentage Assignment using Bi-Allelic Genetic Markers

Version 0.0.1

Description Can be used for paternity and maternity assignment and outperforms
conventional methods where closely related individuals occur in the pool of
possible parents. The method compares the genotypes of offspring with any
combination of potentials parents and scores the number of mismatches of these
individuals at bi-allelic genetic markers (e.g. Single Nucleotide Polymorphisms).
It elaborates on a prior exclusion method based on the Homozygous Opposite Test
(HOT; Huisman 2017 <doi:10.1111/1755-0998.12665>) by introducing the
additional exclusion criterion HIPHOP (Homozygous Identical Parents,
Heterozygous Offspring are Precluded; Cockburn et al., in revision). Potential
parents are excluded if they have more mismatches than can
be expected due to genotyping error and mutation, and thereby one can identify the
true genetic parents and detect situations where one (or both) of the true
parents is not sampled. Package 'hiphop' can deal with (a) the case where there
is contextual information about parentage of the mother (i.e. a female has been
seen to be involved in reproductive tasks such as nest building), but paternity
is unknown (e.g. due to promiscuity), (b) where both parents need to be assigned,
because there is no contextual information on which female laid eggs and which
male fertilized them (e.g. polygynandrous mating system where multiple females
and males deposit young in a common nest, or organisms with external fertilisation
that breed in aggregations).
For details: Cockburn, A., Penalba, J.V.,Jaccoud, D.,Kilian, A., Brouwer, L.,
Double, M.C., Margraf, N., Osmond, H.L., van de Pol, M. and Kruuk, L.E.B.
(in revision). HIPHOP: improved paternity assignment among close relatives
using a simple exclusion method for bi-allelic markers.
Molecular Ecology Resources, DOI to be added upon acceptance.

Depends R (>= 3.6)

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

URL https://groups.google.com/d/forum/r-hiphop

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Martijn van de Pol [aut, cre] (<https://orcid.org/0000-0003-4102-4079>),
        Lyanne Brouwer [aut] (<https://orcid.org/0000-0001-6728-4851>),
        Andrew Cockburn [aut] (<https://orcid.org/0000-0002-8531-3350>)

Maintainer Martijn van de Pol <martijn@myscience.eu>

Repository CRAN

Date/Publication 2020-08-19 09:30:03 UTC

R topics documented:

---

combinations   *List of triads to be ranked for parentage assignment.*

Description

An example list of how the output from the hothiphop function looks like. Listed are all possible triad combinations for 8 offspring from the 2018 cohort This dataframe is also used to generate the vignette. See the help file of the hothiphop function for an explanation of columns.

Usage

```
combinations
```

Format

An object of class data.frame with 24192 rows and 24 columns.

---

genotypes   *Genotypes of individuals scored at different loci*

Description

A dataset containing genotypes of 1407 superb fairy wrens (Malurus cyaneus) from five cohorts (breeding seasons 2014-2018) from the Australian National Botanic Garden. Each individual (rows) is scored at 1376 loci (columns) with the scores meaning: 0: homozygotes at common allele; 1: homozygote at rare allele and 2: heterozygotes; NA: locus could not be scored.

Usage

```
genotypes
```

Format

A data frame with 1407 rows (individuals) and 1376 variables (loci)

Author(s)

Andrew Cockburn, <andrew.cockburn@anu.edu.au>

Source

Cockburn et al. (2020) HIPHOP: improved paternity assignment among close relatives using a simple exclusion method for biallelic markers. Molecular Ecology Resources, in revision.

---

hothiphop   *A function that calculates the genetic mismatches according to the hiphop and hot test*

Description

This function calculates the number genetic mismatches according to the hiphop and hot test for any combination of offspring-potential.dam-potential.sire. The HOT test (Homozygous Opposite Test; Huisman 2017) compares the genotype of an offspring with a potential parent: a mismatch is scored when both the offspring and parent are homozygous, but for different alleles. The HIPHOP test (Homozygous Identical Parents, Heterozygous Offspring are Precluded; Cockburn et al. in revision) compares the genotype of an offspring with both potential parents: a mismatch is scored when the offspring is heterozygous and both parent are homozygous for the same allele The resulting output can next be summarized using the 'topmatch()' function.

Usage

```
hothiphop(ind, gen)
```

Arguments

ind	The input file with individuals, which should contain at least the columns brood, individual, type, social.parent, year.
gen	The input file with genotypes, which should contain the loci-names as column headers and the individual names as row-header and should only contain the values 0, 1, 2 or NA.

Value

a dataframe with all possible offspring-potential.dam-potential.sire combinations and their mismatch scores according to the HOT and HIPHOP test, the number of loci this was based on, and some additional relevant information about the social parents and potential dam and sires

year the year or cohort that is being considered, adults can be potential dam or sire in some years, but no in others

brood an identifier of the brood to which the offspring and adults belong/are associated with

offspring an identifier of the offspring

potential.dam an identifier of the potential dam

potential.sire an identifier of the potential sire

hothiphop.parents the sum of the hiphop and hot.parents mismatch score

hiphop the hiphop mismatch score of the offspring with the potential dam and potential sire, expressed as the number of loci giving mismatches

hot.parents the hot score of the offspring with both the potential dam and sire, expressed as the number of loci giving mismatches

hot.dam the hot score of the offspring with the potential dam, expressed as the number of loci giving mismatches

hot.sire the hot mismatch score of the offspring with the potential sire , expressed as the number of loci giving mismatches

hothiphop.dam the sum of the hot.dam and hiphop mismatch score

hothiphop.sire the sum of the hot.sire and hiphop mismatch score

loci.dyad.dam the number of loci at which both the offspring and dam were not NA

loci.dyad.sire the number of loci at which both the offspring and sire were not NA

loci.triad the number of loci at which the offspring, dam and sire were not NA

offspring.heterozygosity proportion of loci at which the offspring was heterozygous

social.mother.sampled if the social.mother genotypic data is in the genotypes file then equal to 1, else 0

social.father.sampled if the social.father genotypic data is in the genotypes file then equal to 1, else 0

is.dam.social if the potential dam is the social mother then equal to 1, else 0

is.sire.social if the potential sire is the social father then equal to 1, else 0

is.dam.within.group if the potential dam is part of the same group (i.e. associated with the same brood) as the offspring, then equal to 1, else 0

is.sire.within.group if the potential sire is part of the same group (i.e. associated with the same brood) as the offspring, then equal to 1, else 0

social.mother identity of the social mother of the offspring

social.father identity of the social father of the offspring

Author(s)

Martijn van de Pol, <martijn@myscience.eu>

Source

Cockburn et al. (2020) HIPHOP: improved paternity assignment among close relatives using a simple exclusion method for biallelic markers. Molecular Ecology Resources, in revision.

References

Cockburn et al. (2020) HIPHOP: improved paternity assignment among close relatives using a simple exclusion method for biallelic markers. Molecular Ecology Resources, in revision.

Huisman, J. (2017). Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. Molecular ecology resources, 17(5), 1009-1024.

Examples

```
results<-hothiphop(ind=individuals[1:22,], gen=genotypes)
head(results)
best<-topmatch(x=results, ranking="hothiphop.parents")
head(best)
```

---

individuals *List of individuals to be compared for parentage assignment.*

Description

A list of 2527 individuals (superb fairy wrens; Malurus cyaneus)) of which their genetics are to be compared to determine parentage. The dataset consists of 1153 offspring, 469 adult females that are potential dams and 905 adult males that are potential sires. Data is from five cohorts (breeding seasons 2014-2018) from the Australian National Botanic Garden. Note that individuals can occur multiple times in the dataset, as adults can have parentage in multiple years. Also offspring can become adults in future years. Also if an individual is associated with multiple broods as social parent this means there can be multiple records per year. Note that the columns social parent and brood are only used to determine whether a potential dam or sire is the social parent, an extra-group (technically extra-brood) parent or within-group parent that is not the social parent (a subordinate).

Usage

```
individuals
```

Format

A data frame with 2527 rows and 6 variables:

brood an identifier of the brood to which the offspring and adults belong/are associated with

individual an identifier of individual

type denotes whether the individual is an offspring, adult female (potential dam) or adult male (potential sire)

social.parent if the individual is the social parent of the brood then equal to 1, else 0

year the year or cohort that is being considered, adults can be potential dam or sire in some years, but no in others

Author(s)

Andrew Cockburn, <andrew.cockburn@anu.edu.au>

Source

Cockburn et al. (2020) HIPHOP: improved paternity assignment among close relatives using a simple exclusion method for biallelic markers. Molecular Ecology Resources, in revision.

| inspect | *A function that inspects the genotypes of individuals* |

Description

This function allows one to inspect the genotypes of all individuals in your input file to see whether genetic data is available, and if so what the proportion of homozygotes, heterozygotes fraction of missing values is.

Usage

```
inspect(ind, gen)
```

Arguments

ind	The input file with individuals, which should contain at least the columns brood, individual, type, social.parent, year.
gen	The input file with genotypes, which should contain the loci-names as column headers and the individual names as row-header and should only contain the values 0, 1, 2 or NA.

Value

the individuals file with a summary of the genotypes attached. #'

brood an identifier of the brood to which the offspring and adults belong/are associated with

individual an identifier of the offspring

type denotes whether the individual is an offspring, adult female (potential dam) or adult male (potential sire)

social.parent if the individual is the social parent of the brood then equal to 1, else 0

year the year or cohort that is being considered, adults can be potential dam or sire in some years, but no in others

sampled if the individual's genotypic data is in the genotypes file then equal to 1, else 0

homozygote.0 the proportion of loci that was scored as 0 (homozygotes at common allele)

homozygote.1 the proportion of loci that was scored as 1 (homozygotes at rare allele)

homozygote.2 the proportion of loci that was scored as 2 (heterozygote)

missing the proportion of loci that has missing values (NA)

number.loci the number of loci that has genotypic information (i.e. not NA)

Author(s)

Martijn van de Pol, <martijn@myscience.eu>

Examples

```
overview<-inspect(ind=individuals[1:22,], gen=genotypes)
head(overview)
```

---

topmatch | *A function selecting the dams and sires with the fewest genetic mismatches to an offspring*

Description

This function summarizes per offspring the top combinations of dam and sire that have the least genetic mismatches according to the hot (Huisman 2017) and/or hiphop (Cockburn et al. in revision) test criteria. In addition to the top matched combinations the summary always also list the social parent, if not among the top X. The user can choose whether one wants to look for the most likely dam and sire with and without assuming that the social mother (or father) is the genetic parent. Furthermore, one can choose on which test score to rank individuals. For more information and worked examples, see the vignette and Cockburn et al. (in revision).

Usage

```
topmatch(
  x,
  ranking,
  condition = "none",
  thres = 99999,
  top = 3,
  unique = "pair"
)
```

Arguments

x	The file with hot and hiphop scores that is generated by the 'hothiphop()' function.
ranking	This sets the mismatch criterion in which dams and sires are ranked, possibilities include ranking="hothiphop.parents", "hiphop", "hot.parents", hot.dam", "hot.sire", "hothiphop.dam", "hothiphop.sire". In some situations it can be useful to supply two ranking criteria, for example to avoid ties (e.g. ranking=c("hot.dam", "hiphop))
condition	Whether or not one wants to condition on either the social mother (condition="mother") or social father (condition="father"). The default is condition="none" in which case all possible dams and sires are considered as genetic parents.
thres	Sets a threshold value for the number of mismatches, if parents have less mismatches then the variable below.threshold in the output will have a value of 1. If 0>=thres<1, the number of mismatches will be re-expressed as a proportion of all loci sampled (not NA). If -1>thres<0 the mismatch scores are expressed as the number of mismatched divided by the number of loci at which the offspring is heterozygote (for the HIPHOP test) or divided by number of loci at which the offspring is homozygote (for the HOT test). This standardization may be sometimes useful for the recalculating the HIPHOP score to account for the fact that some offspring can be scored at more loci than others for this test. By default thres is set to a large integer (99999).
top	Sets the number of top matches that is shown in the output #' @return a dataframe with for each offspring the top X (default top 3, giving 3 rows for each offspring) dam and sire combinations and their mismatch scores according to the HOT and HIPHOP test, the number of loci this was based on, and some additional relevant information about the social parents and potential dam and sires. In addition to the top X offspring-dam-sire with the fewest mismatches, the scores of the social parents are also always listed (if they were not already in the top X ranking). #'

      year the year or cohort that is being considered, adults can be potential dam or sire in some years, but no in others

      brood an identifier of the brood to which the offspring and adults belong/are associated with

      offspring an identifier of the offspring

rank the ranking among all possible combinations of offspring and potential dams and sires, rank 1 is fewest mismatches according to criterion chosen in the function argument 'ranking'

dam an identifier of the potential dam

sire an identifier of the potential sire

dam.type description of the type of potential dam (social parent, extra-group parent, within-group subordinate)

sire.type description of the type of potential sire (social parent, extra-group parent, within-group subordinate)

hothiphop.parents the sum of the hiphop and hot.parents mismatch score

hiphop the hiphop mismatch score of the offspring with the potential dam and potential sire, expressed as the number of loci giving mismatches

hot.parents the hot score of the offspring with both the potential dam and sire, expressed as the number of loci giving mismatches

hot.dam the hot score of the offspring with the potential dam, expressed as the number of loci giving mismatches

hot.sire the hot mismatch score of the offspring with the potential sire , expressed as the number of loci giving mismatches

hothiphop.dam the sum of the hot.dam and hiphop mismatch score

hothiphop.sire the sum of the hot.sire and hiphop mismatch score

below.threshold if the score chosen in argument 'ranking' is below the chose threshold value, then equals to 1, else 0

threshold the chosen threshold value in argument 'threshold'

loci.dyad.dam the number of loci at which both the offspring and dam were not NA

loci.dyad.sire the number of loci at which both the offspring and sire were not NA

loci.triad the number of loci at which the offspring, dam and sire were not NA

offspring.heterozygosity the proportion of loci at which the offspring was heterozygous

social.mother.sampled if the social.mother genotypic data is in the genotypes file then equal to 1, else 0

social.father.sampled if the social.father genotypic data is in the genotypes file then equal to 1, else 0

social.mother identity of the social mother of the offspring

social.father identity of the social father of the offspring

condition the value of the argument 'condition'

ranking the value of the (first) value in the argument 'ranking'

ranking2 the value of the second value in the argument 'ranking'

`unique` When using the ranking criteria hot.dam or hot.sire the best ranked dam or sire will occupy the entire top 3 and this does not allow comparing among dams. By setting unique to "dam" (or "sire") the output shows the top 3 with the best record for each unique dam (or sire).

Author(s)

Martijn van de Pol, <martijn@myscience.eu>

Source

Cockburn et al. (2020) HIPHOP: improved paternity assignment among close relatives using a simple exclusion method for biallelic markers. Molecular Ecology Resources, in revision.

References

Cockburn et al. (2020) HIPHOP: improved paternity assignment among close relatives using a simple exclusion method for biallelic markers. Molecular Ecology Resources, in revision.

Huisman, J. (2017). Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. Molecular ecology resources, 17(5), 1009-1024.

Examples

```
results<-hothiphop(ind=individuals[1:22,], gen=genotypes)
best<-topmatch(x=results, ranking="hothiphop.parents")
head(best)
```

Index