Package: fingers (via r-universe)

May 9, 2026

Version 0.60-6

Date 2026-05-05

Title Identifying Clusters of Related Individuals

Author Karl W Broman [aut, cre]
(<https://orcid.org/0000-0002-4914-6671>), Laura Plantinga
[ctb]

Maintainer Karl W Broman <broman@wisc.edu>

Description Identifying clusters of related individuals.

Depends R (>= 2.10.1)

Imports stats, graphics, grDevices

License GPL-3

URL https://github.com/kbroman/fingers

Repository https://kbroman.r-universe.dev

Date/Publication 2026-05-09 20:20:19 UTC

RemoteUrl https://github.com/kbroman/fingers

RemoteRef HEAD

RemoteSha 12b2ec8593361aedb154e5295ce75ef856999f1a

Contents

---

aedes *Data on Aedes aegypti*

Description

This is RAPD data for 40 loci typed on a set of 10 full-sibling families, with 15 individuals in each family.

Usage

```
data(aedes)
```

Format

The data is a matrix of 150 rows (the individuals) by 40 columns (the RAPD loci). Each entry is a RAPD phenotype, indicating the presence (1) or absence (0) of a band.

Author(s)

Karl W Broman <broman@wisc.edu>

Source

FINGERS software, WC Black IV, Colorado State University

References

BL Apostol, WC Black IV, BR Miller, P Reiter, BJ Beaty (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. Theor Appl Genet 86:991-1000.

See Also

shiff1, simrapd

Examples

```
data(aedes)
```

`calc.dist` *Calculate simple distance matrix*

Description

Calculate the simple distance matrix, by the proportion of mismatches, for a RAPD data set.

Usage

```
calc.dist(dat)
```

Arguments

`dat`	A matrix of size (n.ind x n.mar) containing RAPD phenotypes, with 1 indicating the presence of a band and 0 indicating absence.

Details

For each pair of individuals, we calculate the proportion of RAPD markers (among those where both individuals have complete data) at which one individual shows a band and the other doesn't.

Value

A symmetric matrix of dimension (n.ind x n.ind), containing the distances between individuals.

Author(s)

Karl W Broman <broman@wisc.edu>

References

BL Apostol, WC Black IV, BR Miller, P Reiter, BJ Beaty (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. Theor Appl Genet 86:991-1000.

See Also

`llrdist`, `fingers`

Examples

```
data(aedes)
d <- calc.dist(aedes)
```

`cluster.stat` *Calculate measure of quality of inferred clusters*

Description

Calculate a score indicating how well two sets of clusters conform.

Usage

```
cluster.stat(fam1,fam2,method=c("all","rand","adj","fm","kb"))
```

Arguments

`fam1`	A list of clusters; each component in the list is one family, containing the indices of the individuals in that family.
`fam2`	A list, just like `fam1`.
`method`	A character string indicating whether to calculate the Rand index, the adjusted Rand index, the Fowlkes and Mallows B index, or Karl Broman's index. If `method=all`, a vector with all four indices is returned.

Details

In the Rand index (Rand 1971), one considers all pairs of individuals, and assigns a 1 to a pair if the individuals are either in the same cluster in both `fam1` and `fam2` or are not in the same cluster in both `fam1` and `fam2`, and assigns a 0 to the pair otherwise, and then takes the sum of these, divided by the number of pairs of individuals.

Karl Broman's index (which we don't recommend, but we implement here in order to allow comparisons to be made) is just like the Rand index, but `fam2` is assumed to be the *true* partition, and the set of all pairs in the same group (by `fam2`) and the set of all pairs in different groups (by `fam2`), are given equal weight.

Let n_{ij} be the number of individuals in group i by partition 1 and group j by partition 2. Let $n_{i.} = \sum_j n_{ij}$ and define $n_{.j}$ similarly.

In the adjusted-Rand index (Hubert and Arabie 1985), ...

In the Fowlkes and Mallows B index (Fowlkes and Mallows 1983), ...

Value

The value of a score for comparing two sets of clusters.

Author(s)

Karl W Broman <broman@wisc.edu>

References

WM Rand (1971) Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66:846-850.

L Hubert and P Arabie (1985) Comparing partitions. Journal of Classification. 2:193-218.

EB Fowlkes and CL Mallows (1983) A method for comparing two hierarchical clusterings. Journal of the American Statistical Association 78:553-584.

BS Everitt, S Landau and M Leese (2001) Cluster analysis, 4th edition. Arnold, London, pp. 181-3.

See Also

fingers, true.fams

Examples

```
data(aedes)
f <- freq(aedes)
co <- cutoff(f)
d <- calc.dist(aedes)
fam <- fingers(d,co,make.plot=TRUE)
tf <- true.fams(aedes)
cluster.stat(fam,tf)
cluster.stat(fam,tf,method="fm")
```

---

comp.fams *Compare two sets of clusters*

Description

Give diagnostic information indicating how well two sets of clusters conform.

Usage

```
comp.fams(fam1,fam2)
```

Arguments

fam1	A list of clusters; each component in the list is one family, containing the indices of the individuals in that family.
fam2	A list, just like fam1.

Value

A list with two components. The first component is a contingency table whose (i,j)th element is the number of individuals in cluster i in fam1 and cluster j in fam2. The second component is a list indicating, for each cluster from fam1, the cluster assignment in fam2.

Author(s)

Karl W Broman <broman@wisc.edu>

See Also

cluster.stat, fingers, true.fams

Examples

```
data(aedes)
f <- freq(aedes)
co <- cutoff(f)
d <- calc.dist(aedes)
fam <- fingers(d,co,make.plot=TRUE)
tf <- true.fams(aedes)
comp.fams(fam,tf)
```

---

cutoff *Calculate cutoff for clustering with RAPD markers*

Description

Calculate the cutoff for hierarchical cluster analysis to infer groups of related individuals with RAPD data.

Usage

```
cutoff(f,method=c("qu","meansib","qs","lr"),value=0.2)
```

Arguments

f	A vector of band allele frequencies for a set of RAPD markers.
method	The method to use to form the cutoff: a quantile of the distribution of distances among unrelated (qu), the mean distance between siblings (meansib), a quantile of the distribution of distances among siblings (qs), or the likelihood ratio for unrelateds vs. siblings (lr).
value	For method="qu" or method="qs", this should specify the quantile; for method="lr", this should specify the likelihood ratio.

Value

The cutoff (a single value).

Author(s)

Karl W Broman <broman@wisc.edu>

References

BL Apostol, WC Black IV, BR Miller, P Reiter, BJ Beaty (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. Theor Appl Genet 86:991-1000.

See Also

cutoff.llr, freq, pull.markers, fingers

Examples

```
data(aedes)
f <- freq(aedes)
co1 <- cutoff(f,method="meansib")
co2 <- cutoff(f,method="qu",value=0.2)
co3 <- cutoff(f,method="qs",value=0.9)
co4 <- cutoff(f,method="lr",value=4.0)
```

---

cutoff.llr *Calculate cutoff for clustering with RAPD markers*

Description

Calculate a cutoff (for the LLR distance measure) for hierarchical cluster analysis to infer groups of related individuals with RAPD data.

Usage

```
cutoff.llr(f,method=c("qu","meansib","qs","lr"),value=0.2)
```

Arguments

f	A vector of band allele frequencies for a set of RAPD markers.
method	The method to use to form the cutoff: a quantile of the distribution of distances among unrelated (qu), the mean distance between siblings (meansib), a quantile of the distribution of distances among siblings (qs), or the likelihood ratio for unrelateds vs. siblings (lr).
value	For method="qu" or method="qs", this should specify the quantile; for method="lr", this should specify the likelihood ratio.

Value

The cutoff (a single value).

Author(s)

Karl W Broman <broman@wisc.edu>

See Also

cutoff, llrdist, freq, pull.markers, fingers

Examples

```
data(aedes)
f <- freq(aedes)
co1 <- cutoff.llr(f,method="meansib")
co2 <- cutoff.llr(f,method="qu",value=0.2)
co3 <- cutoff.llr(f,method="qs",value=0.9)
co4 <- cutoff.llr(f,method="lr",value=4.0)
```

---

dist.image *Plot distance matrix*

Description

Plot the distance matrix for a RAPD data set, with (optionally) lines drawn separating clusters of individuals.

Usage

```
dist.image(dist,fams=NULL,col=topo.colors(1+ncol(dist)),...)
```

Arguments

dist	A matrix of size (n.ind x n.ind), containing the distances between pairs of individuals.
fams	A list of clusters; each component in the list is one inferred family, containing the indices of individuals placed in that family.
col	Colors to use in the plot; see image.
...	Other arguments to pass to image.

Value

The function calls image in order to create an image of the distance matrix.

Author(s)

Karl W Broman <broman@wisc.edu>

See Also

calc.dist, true.fams

Examples

```
data(aedes)
f <- freq(aedes)
co <- cutoff(f)
d <- calc.dist(aedes)
fam <- fingers(d,co,make.plot=TRUE)
dist.image(d,fam)
```

`fingers` *Infer clusters of related individuals*

Description

Perform hierarchical clustering to infer groups of related individuals with RAPD data.

Usage

```
fingers(dist,cutoff=NULL,method=c("average","complete",
         "mcquitty","single","ward"),truefam=NULL,
         make.plot=FALSE,just.plot=FALSE)
```

Arguments

`dist`	A matrix of size (n.ind x n.ind) containing the distances between individuals.
`cutoff`	A value to use to cut off the dendogram formed by hierarchical clustering in order to define a set of clusters. (Optional, but if NULL, the argument `truefam` must be included.)
`method`	A hierarchical clustering method. See `hclust`. *Note:* We haven't allowed `centroid` or `median`, because these weren't working for us.
`truefam`	The true family structure; used only if `cutoff` is NULL, in which case all possible cutoffs are tried, and that giving the maximum adjusted Rand index is used.
`make.plot`	If TRUE, make a plot of the dendogram formed by hierarchical clustering.
`just.plot`	If TRUE, just make the plot; don't return the inferred families. (In this case, the `cutoff` argument is not needed.)

Details

We use the function `hclust` to do the cluster analysis.

Value

A list of clusters; each component in the list is one inferred family, containing the indices of individuals placed in that family. The cutoff used is included as an attribute. Use `attr(result,"cutoff")` to obtain this value.

Author(s)

Karl W Broman <broman@wisc.edu>

References

BL Apostol, WC Black IV, BR Miller, P Reiter, BJ Beaty (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. Theor Appl Genet 86:991-1000.

See Also

cutoff, cutoff.llr, calc.dist, llrdist, cluster.stat, true.fams, freq, pull.markers

Examples

```
data(aedes)
f <- freq(aedes)
co <- cutoff(f)
d <- calc.dist(aedes)
fam <- fingers(d,co,make.plot=TRUE)
tf <- true.fams(aedes)
cluster.stat(fam,tf)
```

---

freq *Estimate RAPD allele frequencies*

Description

Estimate the frequency of the band allele for a set of RAPD markers.

Usage

```
freq(dat)
```

Arguments

dat	A matrix of size (n.ind x n.mar) containing RAPD phenotypes, with 1 indicating the presence of a band and 0 indicating absence.

Details

The RAPDs are assumed to be in Hardy-Weinberg equilibrium, and so the frequency of the band allele is estimated as $\hat{p} = 1 - \sqrt{1 - \bar{x}}$ where $\bar{x}$ is the proportion of individuals showing a band.

Value

A vector of length n.mar, containing the estimated frequencies of the band allele for each RAPD marker.

Author(s)

Karl W Broman <broman@wisc.edu>

References

BL Apostol, WC Black IV, BR Miller, P Reiter, BJ Beaty (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. Theor Appl Genet 86:991-1000.

See Also

pull.markers

Examples

```
data(aedes)
f <- freq(aedes)
```

---

llrdist *Calculate distance matrix based on log likelihood ratio*

Description

Calculate a distance matrix, based on the log likelihood ratio comparing the hypotheses of full sibling versus unrelated, for a RAPD data set.

Usage

```
llrdist(dat,p=freq(dat))
```

Arguments

dat	A matrix of size (n.ind x n.mar) containing RAPD phenotypes, with 1 indicating the presence of a band and 0 indicating absence.
p	A vector of band allele frequencies.

Details

For each pair of individuals, at each locus, we calculate the log likelihood ratio (LLR) comparing the hypotheses *unrelated* with *siblings*, with the data being B (both have band), N (neither have band) or D (one has band, the other doesn't). These LLRs are *averaged* across individuals. **Note**: at each locus, we re-center the LLRs so that the minimum of the LLRs among B/N/D is 0; this makes the resulting distances ≥ 0.

Calculations are performed in a C program.

Value

A symmetric matrix of dimension (n.ind x n.ind), containing the distances between individuals.

Author(s)

Karl W Broman <broman@wisc.edu>

See Also

calc.dist, fingers

Examples

```
data(aedes)
f <- freq(aedes)
dis <- llrdist(aedes,f)
```

---

pull.markers *Extract markers with allele frequencies in specified range*

Description

Extract markers from a RAPD data set that have allele frequencies within a specified range.

Usage

```
pull.markers(dat,lo=0.1,hi=0.6,f=freq(dat))
```

Arguments

dat	A matrix of size (n.ind x n.mar) containing RAPD phenotypes, with 1 indicating the presence of a band and 0 indicating absence.
lo	Lower bound for band allele frequency.
hi	Upper bound for band allele frequency.
f	Vector of band allele frequencies (included in order to avoid recalculating it, if possible).

Value

A matrix, like the argument dat, but containing only those markers with band allele frequency between lo and hi.

Author(s)

Karl W Broman <broman@wisc.edu>

See Also

freq

Examples

```
data(shiff1)
f <- freq(shiff1)
subset <- pull.markers(shiff1, 0.1, 0.6, f)
```

---

shiff1 *Schistosome data*

Description

This is RAPD data for 35 loci typed on a set of 135 individuals.

Usage

```
data(shiff1)
```

Format

The data is a matrix of 135 rows (the individuals) by 35 columns (the RAPD loci). Each entry is a RAPD phenotype, indicating the presence (1) or absence (0) of a band.

Author(s)

Karl W Broman <broman@wisc.edu>

Source

Clive Shiff, Molecular Microbiology and Immunology, Bloomberg School of Public Health, The Johns Hopkins University

See Also

shiff2, shiff3, aedes, simrapd

Examples

```
data(shiff1)
```

shiff2 *Schistosome data*

Description

This is RAPD data for 10 loci typed on a set of 135 individuals. Markers with estimated band allele frequencies outside of the range 0.1-0.6 have been removed.

Usage

```
data(shiff2)
```

Format

The data is a matrix of 135 rows (the individuals) by 10 columns (the RAPD loci). Each entry is a RAPD phenotype, indicating the presence (1) or absence (0) of a band.

Author(s)

Karl W Broman <broman@wisc.edu>

Source

Clive Shiff, Molecular Microbiology and Immunology, Bloomberg School of Public Health, The Johns Hopkins University

See Also

shiff1, shiff3, aedes, simrapd

Examples

```
data(shiff2)
```

shiff3 *Schistosome data*

Description

This is RAPD data for 10 loci typed on a set of 125 individuals. Markers with estimated band allele frequencies outside of the range 0.1-0.6 have been removed. Individuals with one or more missing values have been removed.

Usage

```
data(shiff3)
```

Format

The data is a matrix of 125 rows (the individuals) by 10 columns (the RAPD loci). Each entry is a RAPD phenotype, indicating the presence (1) or absence (0) of a band.

Author(s)

Karl W Broman <broman@wisc.edu>

Source

Clive Shiff, Molecular Microbiology and Immunology, Bloomberg School of Public Health, The Johns Hopkins University

See Also

shiff1, shiff2, aedes, simrapd

Examples

```
data(shiff3)
```

---

simrapd *Simulate RAPD data*

Description

Simulates RAPD data for a set of sibling families.

Usage

```
simrapd(n.sib = rep(15,10), p = c(rep(0.125,8),rep(0.175,5),rep(0.225,5),
           rep(0.275,8),rep(0.325,3),rep(0.375,4),
           rep(0.475,4),rep(0.575,3)))
```

Arguments

n.sib	A vector giving the number of siblings per family (length is the number of families).
p	A vector of frequencies of the band allele at each marker (length is the number of markers).

Details

The RAPDs are assumed to be in Hardy-Weinberg equilibrium.

Value

A matrix of dimension (n.ind x n.mar), giving the RAPD phenotypes for each individual at each marker, with 1 indicating a band and 0 indicating no band.

Author(s)

Karl W Broman <broman@wisc.edu>

References

BL Apostol, WC Black IV, BR Miller, P Reiter, BJ Beaty (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. Theor Appl Genet 86:991-1000.

See Also

simulfams

Examples

```
data <- simrapd(rep(20,5), p=runif(40, 0.1, 0.6))
```

---

simulfams *Simulate RAPD data*

Description

Simulates RAPD data for a set of sibling families.

Usage

```
simulfams(n.sib=sample(5:20,size=sample(5:20,size=1),replace=TRUE),
          p=runif(sample(5:15,size=1),min=0.1,max=0.6))
```

Arguments

n.sib	A vector giving the number of siblings per family (length is the number of families).
p	A vector of frequencies of the band allele at each marker (length is the number of markers).

Details

The RAPDs are assumed to be in Hardy-Weinberg equilibrium.

Value

A matrix of dimension (n.ind x n.mar), giving the RAPD phenotypes for each individual at each marker, with 1 indicating a band and 0 indicating no band.

Author(s)

Laura Plantinga and Karl Broman <broman@wisc.edu>

References

BL Apostol, WC Black IV, BR Miller, P Reiter, BJ Beaty (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. Theor Appl Genet 86:991-1000.

See Also

simrapd

Examples

```
data <- simulfams(rep(20,5), p=runif(40, 0.1, 0.6))
```

`true.fams` *Identify the true clusters*

Description

Use the row names of a RAPD data set to identify the true sets of families.

Usage

```
true.fams(dat)
```

Arguments

`dat`	A matrix of size (n.ind x n.mar) containing RAPD phenotypes, with 1 indicating the presence of a band and 0 indicating absence. The row names (identifying individuals) are assumed to be of the form "family-individual"

Value

A list of clusters; each component in the list is one inferred family, containing the indices of individuals placed in that family.

Author(s)

Karl W Broman <broman@wisc.edu>

References

BL Apostol, WC Black IV, BR Miller, P Reiter, BJ Beaty (1993) Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. Theor Appl Genet 86:991-1000.

See Also

aedes, simrapd, fingers, cluster.stat

Examples

```
data(aedes)
tf <- true.fams(aedes)
```

Index