

Package ‘ipADMIXTURE’

July 22, 2025

Title Iterative Pruning Population Admixture Inference Framework

Version 0.1.2

Description A data clustering package based on admixture ratios (Q matrix) of population structure. The framework is based on iterative Pruning procedure that performs data clustering by splitting a given population into subclusters until meeting the condition of stopping criteria the same as ipPCA, iNJclust, and IPCAPS frameworks. The package also provides a function to retrieve phylogeny tree that construct a neighbor-joining tree based on a similar matrix between clusters. By given multiple Q matrices with varying a number of ancestors (K), the framework define a similar value between clusters i,j as a minimum number K^* that makes majority of members of two clusters are in the different clusters. This K^* reflex a minimum number of ancestors we need to splitting cluster i,j into different clusters if we assign K^* clusters based on maximum admixture ratio of individuals. The publication of this package is at Chainarong Amornbunchornvej, Pongsakorn Wangkumhang, and Sisades Tongsimma (2020) <[doi:10.1101/2020.03.21.001206](https://doi.org/10.1101/2020.03.21.001206)>.

Depends R (>= 3.5.0)

Imports stats,treemap,ape

URL <https://github.com/DarkEyes/ipADMIXTURE>

BugReports <https://github.com/DarkEyes/ipADMIXTURE/issues>

Language en-US

License GPL-3

Encoding UTF-8

LazyData true

Suggests knitr, rmarkdown

VignetteBuilder knitr

RoxygenNote 7.3.2

NeedsCompilation no

Author Chainarong Amornbunchornvej [aut, cre] (ORCID:
<<https://orcid.org/0000-0003-3131-0370>>)

Maintainer Chainarong Amornbunchornvej <grandca@gmail.com>

Repository CRAN

Date/Publication 2025-05-06 09:50:01 UTC

Contents

biclustFunc	2
getPhyloTree	3
human27pop_labels	4
human27pop_Qmat	4
ipADMIXTURE	5
plotAdmixClusters	6
plotClusterLeaves	7
printClustersFromLabels	7
UD1labels	8
UD1_Qmat	8
Index	9

biclustFunc	<i>biclustFunc function</i>
-------------	-----------------------------

Description

biclustFunc is a binary clustering function using hierarchical clustering.

Usage

```
biclustFunc(Qmat, admixRatioThs = 0.5, method = "average")
```

Arguments

Qmat	is a Q matrix that contains admixture ratios of all individuals where the $Qmat[i, j]$ represents the admixture ratio of ancestor j for individual i .
admixRatioThs	is a threshold to determine that if a cluster has <code>maxDiffAdmixRatio</code> lower than threshold, then the cluster is a homogeneous cluster.
method	is a method parameter of <code>hclust</code> object for hierarchical clustering analysis. The default is "average".

Value

This function returns binary clustering results.

heteroFlag	is a flag that represents a status whether a given cluster is heterogeneous (having sub-clusters). It is TRUE if <code>maxDiffAdmixRatio</code> \geq <code>admixRatioThs</code> .
clusterInx	is a vector of clustering assignment where <code>indexClsVec[i]</code> is a cluster number of individual i .
meanDiffAdmixRatio	is a vector of magnitude-difference of admixture ratios. It is calculated by splitting a given cluster into two sub-clusters. Then, we take the absolute on the difference between mean admixture ratios of sub-clusters.

Qmat1	is a Q matrix of sub-cluster #1 after splitting a given cluster into two sub-clusters that contains admixture ratios of all individuals where the $Qmat[i, j]$ represents the admixture ratio of ancestor j for individual i.
Qmat2	is a Q matrix of sub-cluster #2 after splitting a given cluster into two sub-clusters that contains admixture ratios of all individuals where the $Qmat[i, j]$ represents the admixture ratio of ancestor j for individual i.
maxDiffAdmixRatio	is a maximum of magnitude-difference of admixture ratios for a given cluster before splitting into two sub-clusters.

Examples

```
# Running biclustFunc on Q matrix of 27 human population dataset where K = 12
obj<-biclustFunc(Qmat=ipADMIXTURE::human27pop_Qmat[[11]], admixRatioThs =0.15)
```

getPhyloTree	<i>getPhyloTree</i>
--------------	---------------------

Description

getPhyloTree is function that reports a phylogenetic tree of clusters based on admixture analysis. The phylogeny tree that construct a neighbor-joining tree based on a similar matrix between clusters. By given multiple Q matrices with varying a number of ancestors (K), the framework define a similar value between clusters i,j as a minimum number K that makes majority of members of two clusters are in the different ancestor groups. This K reflexes a minimum number of ancestors we need to splitting cluster i,j into different clusters if we assign K clusters based on maximum admixture ratio of individuals.

Usage

```
getPhyloTree(QmatList, indexClsVec)
```

Arguments

QmatList	is list of Q matrix where $QmatList[[k]]$ is a Q matrix with k+1 ancestors.
indexClsVec	is a vector of clustering assignment where $indexClsVec[i]$ is a cluster number of individual i.

Value

This function returns an object of nj tree as well as a matrix minDiffAncestorClsMat that is used as a similarity matrix.

tree	is an object of nj tree calculated by <code>ape::nj()</code> function on a dissimilarity version of minDiffAncestorClsMat.
------	--

minDiffAncestorClsMat

is a minimum-ancestor-number matrix in the group level where `minDiffAncestorClsMat[i, j]` is a minimum number of ancestors that make `i` and `j` to be different ancestor groups while `minDiffAncestorClsMat[i, j]-1` makes majority of members from `i` and `j` belong to the same ancestor group.

minDiffAncestorMat

is a minimum-ancestor-number matrix in the individual level where `minDiffAncestorMat[i, j]` is a minimum number of ancestors that make `i` and `j` to be different ancestor groups

Examples

```
# Running ipADMIXTURE on Q matrices (K=2-12) of 27 human population dataset.
h27pop_obj<-ipADMIXTURE(Qmat=ipADMIXTURE::human27pop_Qmat[[11]], admixRatioThs =0.15)
out<-ipADMIXTURE::getPhyloTree(ipADMIXTURE::human27pop_Qmat,h27pop_obj$indexClsVec)
plot(out$tree)
```

human27pop_labels *Labels of 27 human populations*

Description

Labels of 27 human populations

Usage

human27pop_labels

Format

Labels of 27 human populations. :

human27pop_labels It is a vector of labels of 544 individuals. There are 27 populations. ...

human27pop_Qmat *A list of Q matrices of 27 human populations*

Description

A dataset containing admixture ratios of 544 individuals from 27 human populations where the number of ancestors ranges from 2 to 12. This dataset was the result of running ADMIXTURE software developed by Zhou, H., et al. (2011). A quasi-Newton acceleration for high-dimensional optimization algorithms. *Statistics and computing*, 21(2), 261-273. on the 27-human-population dataset published by Xing, J., Watkins, W. S. et al. (2009). Fine-scaled human genetic structure revealed by SNP microarrays. *Genome research*, 19(5), 815-825.

Usage

```
human27pop_Qmat
```

Format

A list of Q matrices of 544 individuals from 27 human populations. There are 2-12 ancestors in the list.

human27pop_Qmat It is list of Q matrices that contains admixture ratios of 544 individuals from the 27 population human dataset. `human27pop_Qmat[[k]][i, j]` is the admixture ratio of *j*th ancestor for *i*th individual in the (*k*+1)-ancestor Q matrix. ...

ipADMIXTURE

Iterative Pruning Population Admixture Inference Framework (ipADMIXTURE)

Description

A data clustering package based on admixture ratios (Q matrix) of population structure.

The framework is based on iterative Pruning procedure that performs data clustering by splitting a given population into subclusters until meeting the condition of stopping criteria the same as ipPCA, iNJclust, and IPCAPS frameworks. The package also provides a function to retrieve phylogeny tree that construct a neighbor-joining tree based on a similar matrix between clusters. By given multiple Q matrices with varying a number of ancestors (K), the framework define a similar value between clusters *i, j* as a minimum number K that makes majority of members of two clusters are in the different clusters. This K reflexes a minimum number of ancestors we need to splitting cluster *i, j* into different clusters if we assign K clusters based on maximum admixture ratio of individuals.

Usage

```
ipADMIXTURE(Qmat, admixRatioThs, method = "average")
```

Arguments

Qmat	is a Q matrix that contains admixture ratios of all individuals where the <code>Qmat[i, j]</code> represents the admixture ratio of ancestor <i>j</i> for individual <i>i</i> .
admixRatioThs	is a threshold to determine that if a cluster has <code>maxDiffAdmixRatio</code> lower than threshold, then the cluster is a homogeneous cluster.
method	is a method parameter of <code>hclust</code> object for hierarchical clustering analysis. The default is "average".

Value

This function returns clustering results in a form of an object of ipADMIXTURE class. The object contains the following items.

<code>indexClsVec</code>	is a vector of clustering assignment where <code>indexClsVec[i]</code> is a cluster number of individual <code>i</code> .
<code>homoClusters</code>	is a list of cluster objects where each object contains member indices, cluster's <code>maxDiffAdmixRatio</code> , ID, etc.
<code>maxDiffAdmixRatioVec</code>	is a vector of <code>maxDiffAdmixRatios</code> for all clusters.
<code>Qmat</code>	is a Q matrix that contains admixture ratios of all individuals where the <code>Qmat[i, j]</code> represents the admixture ratio of ancestor <code>j</code> for individual <code>i</code> .
<code>admixtureRatioThs</code>	is a threshold to determine that if a cluster has <code>maxDiffAdmixRatio</code> lower than threshold, then the cluster is a homogeneous cluster.

Author(s)

Chainarong Amornbunchornvej, <chai@ieee.org>

Examples

```
# Running ipADMIXTURE on Q matrix of 27 human population dataset where K = 12
h27pop_obj<-ipADMIXTURE(Qmat=ipADMIXTURE::human27pop_Qmat[[11]], admixtureRatioThs =0.15)
```

`plotAdmixClusters` *plotAdmixClusters*

Description

`plotAdmixClusters` is function that plots admixture ratios where the x axis represents individuals with cluster labels and y axis represents admixture ratios.

Usage

```
plotAdmixClusters(obj)
```

Arguments

`obj` is an object of ipADMIXTURE class.

Examples

```
h27pop_obj<-ipADMIXTURE(Qmat=ipADMIXTURE::human27pop_Qmat[[11]], admixtureRatioThs =0.15)
ipADMIXTURE::plotAdmixClusters(h27pop_obj)
```

plotClusterLeaves *plotClusterLeaves*

Description

plotClusterLeaves is function that plots clusters in a form of treemap plot. Subsquares represent clusters. Each subsquare contains cluster label (ID), number of members (N), and a maximum of manitude-difference of admixture ratios (md). A size of each subsquare represents a ratio of member numbers compared to other clusters. A color represents an md value of cluster.

Usage

```
plotClusterLeaves(obj)
```

Arguments

obj is an object of ipADMIXTURE class.

Examples

```
h27pop_obj<-ipADMIXTURE(Qmat=ipADMIXTURE::human27pop_Qmat[[11]], admixRatioThs =0.15)
ipADMIXTURE::plotClusterLeaves(h27pop_obj)
```

printClustersFromLabels
 printClustersFromLabels

Description

printClustersFromLabels is function that reports that clustering results in text mode.

Usage

```
printClustersFromLabels(obj, labels)
```

Arguments

obj is an object of ipADMIXTURE class.
labels is a vector of labels of all individuals.

Examples

```
h27pop_obj<-ipADMIXTURE(Qmat=ipADMIXTURE::human27pop_Qmat[[11]], admixRatioThs =0.15)
ipADMIXTURE::printClustersFromLabels(h27pop_obj, ipADMIXTURE::human27pop_labels)
```

UD1labels

Labels of 20 simulation populations

Description

Labels of 20 simulation populations

Usage

UD1labels

Format

Labels of 20 populations. :

UD1labels It is a vector of labels of 1200 individuals. There are 20 populations. ...

UD1_Qmat

A list of Q matrices of simulation of 20 populations

Description

A dataset containing admixture ratios of 1200 individuals from 20 simulation populations where the number of ancestors ranges from 2 to 18. This dataset was the result of running LEA library developed by Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925-929. on the 20-simulation-population dataset published by Limpiti, T., et al. (2014). iNJclust: iterative neighbor-joining tree clustering framework for inferring population structure. *IEEE/ACM transactions on computational biology and bioinformatics*, 11(5), 903-914.

Usage

UD1_Qmat

Format

A list of Q matrices of 1200 individuals from 20 populations. There are Q matrices that have the number of ancestors ranges from from 2 to 18.

UD1_Qmat It is list of Q matrices that contains admixture ratios of 1200 individuals from the 20-population dataset. $UD1_Qmat[[k]][i, j]$ is the admixture ratio of jth ancestor for ith individual in the (k+1)-ancestor Q matrix. ...

Index

* datasets

human27pop_labels, 4

human27pop_Qmat, 4

UD1_Qmat, 8

UD1labels, 8

biclustFunc, 2

getPhyloTree, 3

human27pop_labels, 4

human27pop_Qmat, 4

ipADMIXTURE, 5

plotAdmixClusters, 6

plotClusterLeaves, 7

printClustersFromLabels, 7

UD1_Qmat, 8

UD1labels, 8