# Package 'EILA'

February 19, 2015

**Type** Package

**Title** Efficient Inference of Local Ancestry

**Version** 0.1-2

**Date** 2013-09-09

**Author** James J. Yang, Jia Li, Anne Buu, and L. Keoki Williams

**Maintainer** James J. Yang <jyangstat@gmail.com>

**Description** Implementation of Efficient Inference of Local Ancestry
using fused quantile regression and k-means classifier

**Depends** R (>= 2.10), class, quantreg

**License** GPL (>= 2)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2013-09-14 07:48:33

## R topics documented:

## `ceuchdyri` *The CEU-CHD-YRI admixed simulation data*

### Description

The list data has five compoents: `admixed`, `anc1`, `anc2`, `anc3`, `true.local.ancestry`, `position`. The data matrixes were simulated data using the allele frequencies from HapMap Population CEU (Utah residents with Northern and Western European ancestry from the CEPH collection), CHD (Chinese in Metropolitan Denver, Colorado), and YRI (Yoruba in Ibadan, Nigeria (West Africa)).

## Usage

```
ceuchdyri
```

## Format

A list consisting of `admixed` (10000 rows, 30 columns), `anc1` (10000 rows, 60 columns), `anc2` (10000 rows, 60 columns), `anc3` (10000 rows and 60 columns), `true.local.ancestry` (10000 rows, 60 columns), and `position` (a vector of length 10000).

The elements of the matrixes (`admixed`, `anc1`, `anc2`, `anc3`) are the genotypes being coded as the number of copies of the variant allele present, 0, 1 or 2. The elements of matrix true.local.ancestry are the true local ancestries being coded as alleles identification from ancestries 1, 2, or 3. The physical positions (in base pairs unit) are in position.

## Source

none

## References

Yang, J. J., LI, J., Buu, A., and Williams, L. K. (2013) Efficient Inference of Local Ancestry. Bioinformatics. doi: 10.1093/bioinformatics/btt488

## Examples

```
data(ceuchdyri)
str(ceuchdyri)
```

---

# `ceuyri` *The CEU-YRI admixed simulation data*

---

## Description

The list data has five compoents: `admixed`, `anc1`, `anc2`, `true.local.ancestry`, `position`. The data matrixes were simulated data using the allele frequencies from HapMap Population CEU (Utah residents with Northern and Western European ancestry from the CEPH collection) and YRI (Yoruba in Ibadan, Nigeria (West Africa)).

## Usage

```
ceuyri
```

## Format

A list consisting of admixed (2300 rows, 30 columns), anc1 (2300 rows, 60 columns), anc2 (2300 rows, 60 columns), true.local.ancestry (2300 rows, 60 columns), and position (a vector of length 2300).

The elements of the matrixes (admixed, anc1, and2) are the genotypes being coded as the number of copies of the variant allele present, 0, 1 or 2. The elements of matrix true.local.ancestry are the true local ancestries being coded as alleles identification from ancestries 1 or 2. The physical positions (in base pairs unit) are in position.

## Source

none

## References

Yang, J. J., LI, J., Buu, A., and Williams, L. K. (2013) Efficient Inference of Local Ancestry. Bioinformatics. doi: 10.1093/bioinformatics/btt488

## Examples

```
data(ceuyri)
str(ceuyri)
```

# eila *A function to infer local ancestry*

## Description

A function to infer local ancestry, using fused quantile regression and $k$-means classifier.

## Usage

```
eila(admixed, position, anc1, anc2, anc3 = NULL, lambda = 15, rng.seed = 172719943)
```

## Arguments

admixed A SNP matrix of admixted individuals with SNPs in the rows, samples in the columns. The elements of the matrix are the genotypes being coded as the number of copies of the variant allele present, 0, 1 or 2.

position A vector for the physical postions of the SNPs

anc1 A SNP matrix of ancestry 1 samples with SNPs in the rows, samples in the columns. The elements of the matrix are the genotypes being coded as the number of copies of the variant allele present, 0, 1 or 2.

anc2 A SNP matrix of ancestry 2 samples with SNPs in the rows, samples in the columns. The elements of the matrix are the genotypes being coded as the number of copies of the variant allele present, 0, 1 or 2.

`anc3` An optional SNP matrix of ancestry 3 samples with SNPs in the rows, samples in the columns (default NULL). The elements of the matrix are the genotypes being coded as the number of copies of the variant allele present, 0, 1 or 2.

`lambda` A number controlling the smoothness of the fused quantile regression (default 15).

`rng.seed` The seed used for the random number generator (default 172719943) for reproducibility of simulated equally admixed individuals.

## Details

`eila` is an function for inferring local ancestry.

The admixed samples are assumed as descended from ancestry 1 ancestry 2, or ancestry 3. The data matrixes of admixed samples and ancestral samples are coded as thee number of copies of the variant allele present (0, 1, or 2). The physical positions of SNPs are in base pairs unit.

The method for efficient inference of local ancestry (EILA) in admixed individuals is based on three steps. The first step assigns a numerical score (with a range of 0 to 1) to genotypes in admixed individuals in order to better quantify the closeness of the SNPs to a certain ancestral population. The second step uses fused quantile regression to identify breakpoints of the ancestral haplotypes. In the third step, the $k$-means classifier is used to infer ancestry at each locus.

The major strength of EILA is that it relaxes the assumption of linkage equilibrium and uses all genotyped SNPs rather than only unlinked loci to increase the power of inference. Another important strength of this method is its higher accuracy and lower variation.

## Value

`local.ancestry` The inferred local ancestry matrix with the same dimensions of `admixed`. The elements of the matrix are local ancestry being coded as alleles identification from ancestries 1, 2, or 3. For example, element 12 means one allele is from ancestry 1 and the other allele is from ancestry 2.

`rng.seed` The seed used to call `set.seed` for reproducibility.

`rng.state` Prior to the call to `set.seed`, `rng.state` is the value of `.Random.seed` should `.Random.seed` exist. Otherwise, NULL is returned.

## Author(s)

James J. Yang, Jia Li, Anne Buu, and L. Keoki Williams

## References

Yang, J. J., LI, J., Buu, A., and Williams, L. K. (2013) Efficient Inference of Local Ancestry. Bioinformatics. doi: 10.1093/bioinformatics/btt488

## See Also

set.seed

### Examples

```
## Two ancestries
data(ceuyri)
res.eila <- eila(admixed  = ceuyri$admixed,
                 position = ceuyri$position,
                 anc1     = ceuyri$anc1,
                 anc2     = ceuyri$anc2)
cat("Overall accuracy:", mean((res.eila$local.ancestry ==
                  ceuyri$true.local.ancestry), na.rm=TRUE),"\n")

## Three ancestries
## Not run:
data(ceuchdyri)
res.eila <- eila(admixed  = ceuchdyri$admixed,
                 position = ceuchdyri$position,
                 anc1     = ceuchdyri$anc1,
                 anc2     = ceuchdyri$anc2,
                 anc3     = ceuchdyri$anc3)
cat("Overall accuracy:", mean(res.eila$local.ancestry ==
           ceuchdyri$true.local.ancestry),"\n")

## End(Not run)
```

# Index