Package: malaria.em (via r-universe)

February 5, 2025

Type Package

Title EM Estimation of Malaria Haplotype Probabilities from Multiply
Infected Human Blood Samples

Version 2.0

Date 2012-03-06

Author Xiaohong Li

Maintainer Xiaohong Li <xli@bg-medicine.com>

Description Using EM algorithm to estimate malaria haplotype
probabilities and the number of infections from multiply
infected human blood samples. Estimated haplotype probabilities
and their standard error are reported.

License GPL-2

LazyLoad yes

LazyData yes

Date/Publication 2012-03-06 17:44:45

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

RemoteUrl https://github.com/PlasmoGenEpi/malaria.em

RemoteRef HEAD

RemoteSha 8e020ad607dee0e4e6acc7653688a7ff01800b69

Contents

---

malaria.em   *EM Estimation of Malaria Haplotype Probabilities and Number of In-*
                      fections from Multiply Infected Human Blood Samples

---

Description

Compute the maximum likelihood estimates for malaria haplotype probabilities and number of
infections based on the malaria genotype data using Expectation-Maximization approach.

Usage

```
malaria.em(geno, sizes=c(2), locus.label=NA)
```

Arguments

geno   matrix of genotypes. Each column represents the alleles observed in each of the
                locus. The observed alleles in each locus are separated by a space, and the order
                of columns corresponds to the order of loci on a chromosome. If there are K
                loci, then ncol(geno) = K. Rows represent the genotypes for each subject.

sizes   An integer or a vector of possible number of parasite strains within individu-
                als in observed data. For example, sizes=c(1:6) means the possible number of
                parasite strains is ranging from 1 to 6. If the length of sizes is greater than 1,
                the estimation will assume zero truncated Poisson distribution on the number of
                parasite strains.

locus.label   vector of labels for loci.

Details

This program is an extension of haplo.em() function in haplo.stats package, which computes the
maximum likelihood estimates (MLEs) of haplotype probabilities in diploid population. Since in
malaria setting, the number of the malaria parasite strains is unknown due to the possible multiple
infections, we impose a probability distribution on this number. Parameter sizes determines the
underlying probability distribution. Length of sizes more than one will lead to the estimation as-
suming a truncated Poisson distribution. In this case, the estimation of the Poisson rate and number
of infections will be reported. Otherwise, if sizes is defined as an integer- C for example, the esti-
mation will assume the number of strains is fixed at C. Although this assumption is not applicable
for malaria setting, it is very useful for any genetic data, in which the number of chromosomes is
same across all samples. By default, sizes=2, which reduces to haplo.em() and give the MLEs of
the haplotype probabilities in diploid population.

Value

haplo.prob.tab matrix of unique haplotypes, MLEs of estimated haplotype probabilities, and
                their standard errors.

haplotype   matrix of unique haplotypes. Each row represents a unique haplotype, and the
                number of columns is the number of loci.

haplo.prob	vector of MLEs of haplotype probabilities. The ith element of hap.prob corresponds to the ith row of haplotype.
haplo.prob.std	standard error of the estimated haplotype frequencies.
lambda	estimated Poisson parameter.
NumofInfection	estimated number of infections.
haplo.sets	List of all possible haplotype combinations and their posterior probability per subject. The first column named ids is a vector for row index of subjects after expanding to all possible haplotype combinations for each subject. If ids=i, then i is the ith row of geno. If the ith subject has n possible haplotype combinations that correspond to their marker genotype, then i is repeated n times. The value in the second column is the row numbers of the unique haplotypes in the returned haplotype matrix.
n.haplo.set	vector of maximum number of haplotype combinations per subject that are consistent with their marker data in the matrix geno. The length of n.haplo.set = nrow(geno).
pred.haplo.set	Predicted haplotype combination that is consistent with their marker data for each subject. The values in pred.haplo.set are the row numbers of the unique haplotypes in the returned haplotype matrix.

Author(s)

Xiaohong Li

References

Li, X., Foulkes, A.S., Yucel, R. and Rich, S.M. (2007) An expectation maximization approach to estimate malaria haplotype frequencies in multiply infected children, Statistical Applications in Genetics and Molecular Biology, Vol. 6 : Iss. 1, Article 33.

Examples

```
data(geno)

sizes<-c(2)
ret1<-malaria.em(geno,sizes,locus.label=c("DQB","DRB") )

sizes<-c(1:3)
ret2<-malaria.em(geno,sizes,locus.label=c("DQB","DRB") )
```

Index