

iMKT Pipeline

Brief intro about why we developed the package. The importance of detecting adaptation, and the many MKT derived methodologies. Take from paper.

Loading the package and checking test data

First of all, install (if this is not done yet) and load the package. Notice that iMKT package includes two sample dataframes named mydafdata and mydivergencedata which are the ones used in this tutorial. This makes easier to replicate the vignettes to understand better all the package functionalities.

```
## Load package
#devtools::install_github("sergihervas/iMKT")
library(iMKT)

## Sample daf data
head(mydafdata)
#>      daf      Pi      P0
#> 1 0.025 22490 17189
#> 2 0.075 3217 4780
#> 3 0.125 1616 2874
#> 4 0.175 999 2088
#> 5 0.225 754 1685
#> 6 0.275 679 1443

## Sample divergence data
mydivergencedata
#>      mi      Di      m0      D0
#> 1 2598805 54641 620019 52537
```

The package includes several functions, classified as follows:

- MKT derived methodologies calculation
 - standard(): Standard MKT
 - FWW(): FWW correction
 - DGRP(): DGRP correction
 - asymptoticMK(): Asymptotic MKT
 - iMK(): integrative MKT
 - completeMKT(): perform all previous tests
- iMK using PopFly and PopHuman data
 - loadPopFly(): load PopFlyData
 - loadPopHuman(): load PopHumanData
 - PopFlyAnalysis(): perform any test using PopFlyData
 - PopHumanAnalysis(): perform any test using PopHumanData


```

#>                alpha.symbol Fishers exact test P-value
#> Cutoff = 0      0.2364499      1.480943e-183
#> Cutoff = 0.05   0.5409548      0.000000e+00
#> Cutoff = 0.1    0.5798139      0.000000e+00
#>
#> `$Divergence metrics`
#> `$Divergence metrics`$`Global metrics`
#>      Ka      Ks      omega
#> 1 0.02102543 0.0847345 0.2481331
#>
#> `$Divergence metrics`$`Estimates by cutoff`
#>      omegaA.symbol omegaD.symbol
#> Cutoff = 0      0.05867104      0.1894620
#> Cutoff = 0.05   0.13422877      0.1139043
#> Cutoff = 0.1    0.14387102      0.1042621
#>
#>
#> `$MKT tables`
#> `$MKT tables`$`Cutoff = 0`
#>
#>
#> Table: cutoff
#>
#>      Polymorphism  Divergence
#> -----
#> Neutral class      45101      52537
#> Selected class     35816      54641
#>
#> `$MKT tables`$`Cutoff = 0.05`
#>
#>
#> Table: cutoff
#>
#>      Polymorphism  Divergence
#> -----
#> Neutral class      27912      52537
#> Selected class     13326      54641
#>
#> `$MKT tables`$`Cutoff = 0.1`
#>
#>
#> Table: cutoff
#>
#>      Polymorphism  Divergence
#> -----
#> Neutral class      23132      52537
#> Selected class     10109      54641

```

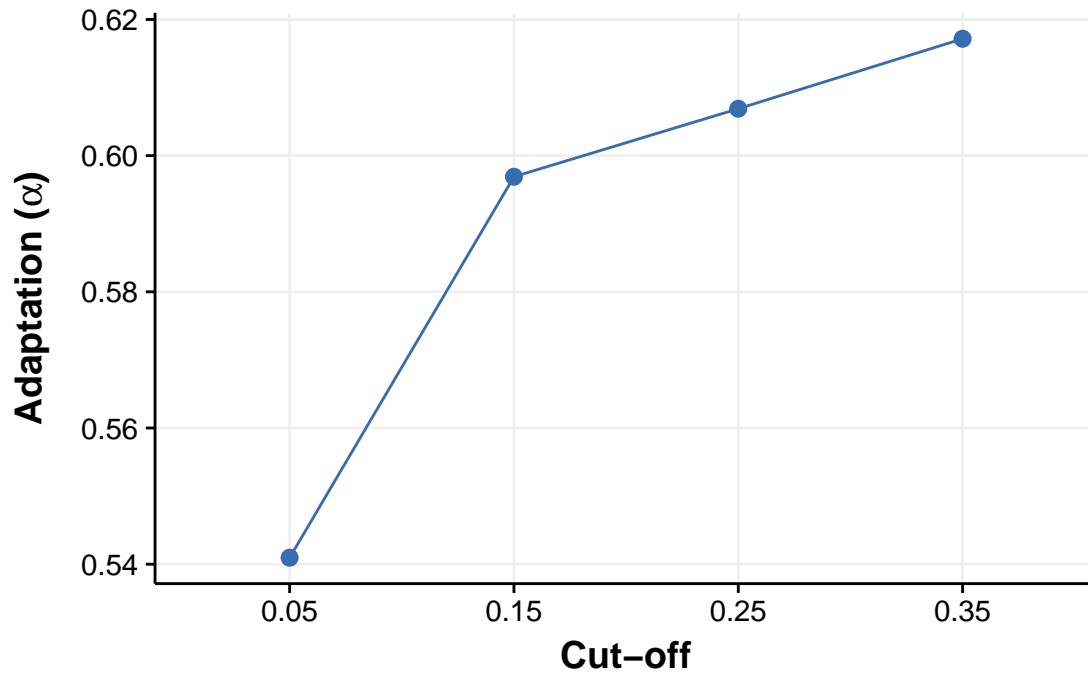
By default the argument **list_cutoff** uses a list of cutoffs with the following values: 0, 0.05, 0.1, and the argument **plot** is set to **FALSE**. This parameters can be customized, like in the following example:

```
FWW(daf=mydafdata, divergence=mydivergencedata, list_cutoff=c(0.05, 0.15,0.25,0.35), plot=TRUE)
```

```

#> $Results
#>               alpha.symbol Fishers exact test P-value
#> Cutoff = 0.05    0.5409548                      0
#> Cutoff = 0.15    0.5969015                      0
#> Cutoff = 0.25    0.6068868                      0
#> Cutoff = 0.35    0.6171609                      0
#>
#> $Graph

```



```

#>
#> $`Divergence metrics`
#> $`Divergence metrics`$`Global metrics`
#>      Ka      Ks      omega
#> 1 0.02102543 0.0847345 0.2481331
#>
#> $`Divergence metrics`$`Estimates by cutoff`
#>      omegaA.symbol omegaD.symbol
#> Cutoff = 0.05      0.1342288    0.11390431
#> Cutoff = 0.15      0.1481110    0.10002208
#> Cutoff = 0.25      0.1505887    0.09754438
#> Cutoff = 0.35      0.1531380    0.09499504
#>
#>
#> $`MKT tables`
#> $`MKT tables`$`Cutoff = 0.05`
#>
#>
#> Table: cutoff
#>
#>      Polymorphism  Divergence
#> -----
#> Neutral class      27912      52537
#> Selected class      13326      54641

```

```

#>
#> $`MKT tables`$`Cutoff = 0.15`
#>
#>
#> Table: cutoff
#>
#>
#>      Polymorphism  Divergence
#> -----
#> Neutral class      20258      52537
#> Selected class      8493      54641
#>
#> $`MKT tables`$`Cutoff = 0.25`
#>
#>
#> Table: cutoff
#>
#>
#>      Polymorphism  Divergence
#> -----
#> Neutral class      16485      52537
#> Selected class      6740      54641
#>
#> $`MKT tables`$`Cutoff = 0.35`
#>
#>
#> Table: cutoff
#>
#>
#>      Polymorphism  Divergence
#> -----
#> Neutral class      13778      52537
#> Selected class      5486      54641

```

DGRP correction

```

DGRP(daf=mydafdata, divergence=mydivergencedata)
#> $Results
#>      alpha.symbol Fishers exact test P-value
#> Cutoff = 0      0.2364499      1.480943e-183
#> Cutoff = 0.05    0.4249071      0.000000e+00
#> Cutoff = 0.2     0.3842950      0.000000e+00
#>
#> $`Divergence metrics`
#> $`Divergence metrics`$`Global metrics`
#>      Ka      Ks      omega
#> 1 0.02102543 0.0847345 0.2481331
#>
#> $`Divergence metrics`$`Estimates by cutoff`
#>      omegaA.symbol omegaD.symbol
#> Cutoff = 0      0.05867104      0.1894620
#> Cutoff = 0.05    0.10543351      0.1426996
#> Cutoff = 0.2     0.09535630      0.1527768
#>

```

```

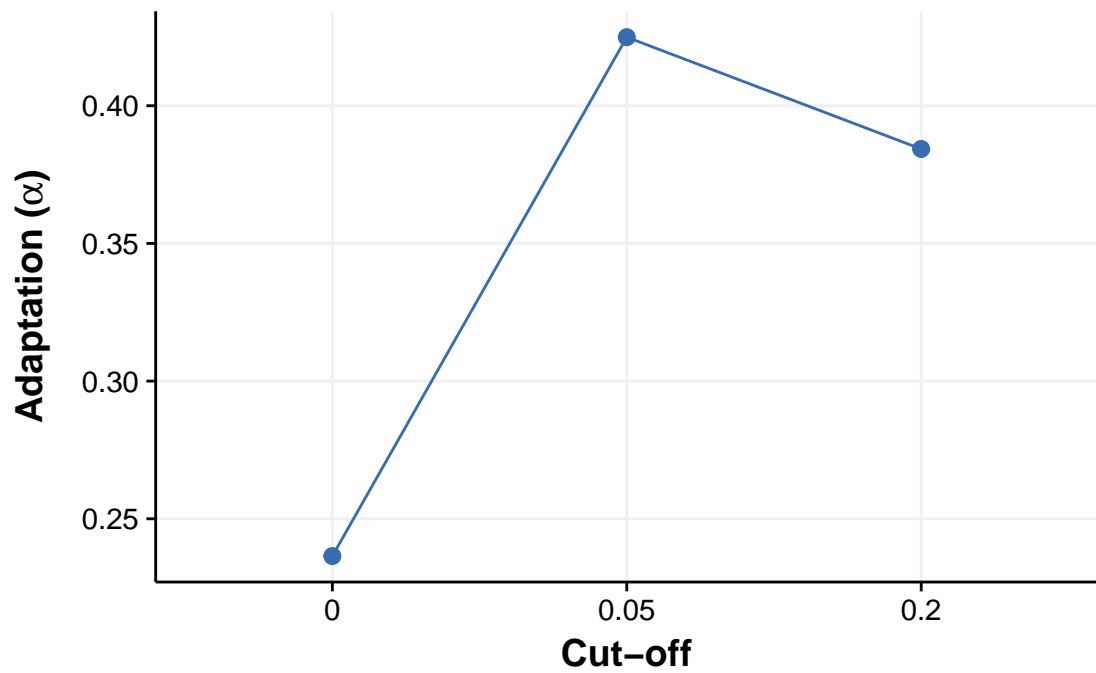
#>
#> $`MKT tables`
#> $`MKT tables`$`Number of segregating sites by DAF category - Cutoff = 0`
#>
#>
#> Table: cutoff
#>
#>
#>      DAF.below.cutoff  DAF.above.cutoff
#> -----
#> Neutral class          0             45101
#> Selected class         0             35816
#>
#> $`MKT tables`$`Number of segregating sites by DAF category - Cutoff = 0.05`
#>
#>
#> Table: cutoff
#>
#>
#>      DAF.below.cutoff  DAF.above.cutoff
#> -----
#> Neutral class        17189             27912
#> Selected class       22490             13326
#>
#> $`MKT tables`$`Number of segregating sites by DAF category - Cutoff = 0.2`
#>
#>
#> Table: cutoff
#>
#>
#>      DAF.below.cutoff  DAF.above.cutoff
#> -----
#> Neutral class        26931             18170
#> Selected class       28322             7494
#>
#> $`MKT tables`$`MKT standard table`
#>
#>
#>
#>      Polymorphism  Divergence
#> -----
#> Neutral class     45101       52537
#> Selected class     35816       54641
#>
#>
#> $Fractions
#>      0      0.05      0.2
#> d 0.810538 0.81053943 0.81053627
#> f 0.189462 0.14269958 0.15277678
#> b 0.000000 0.04676099 0.03668695

DGRP(daf=mydafdata, divergence=mydivergencedata, list_cutoff=c(0,0.05,0.2), plot=TRUE)
#> $Results
#>
#>      alpha.symbol Fishers exact test P-value
#> Cutoff = 0      0.2364499      1.480943e-183
#> Cutoff = 0.05   0.4249071      0.000000e+00
#> Cutoff = 0.2    0.3842950      0.000000e+00

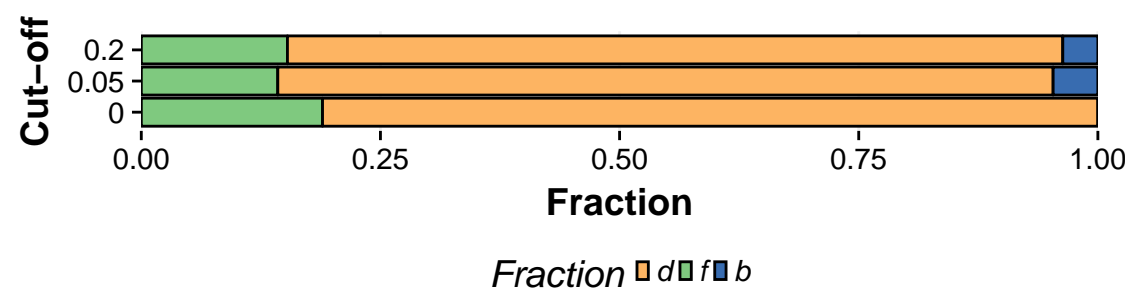
```

```
#>
#> $Graph
```

A



B



```
#>
#> $`Divergence metrics`
#> $`Divergence metrics`$`Global metrics`
#>      Ka      Ks      omega
#> 1 0.02102543 0.0847345 0.2481331
#>
#> $`Divergence metrics`$`Estimates by cutoff`
#>      omegaA.symbol omegaD.symbol
#> Cutoff = 0      0.05867104      0.1894620
#> Cutoff = 0.05    0.10543351      0.1426996
#> Cutoff = 0.2     0.09535630      0.1527768
#>
#>
#> $`MKT tables`
#> $`MKT tables`$`Number of segregating sites by DAF category - Cutoff = 0`
#>
#>
```

```

#> Table: cutoff
#>
#>
#>      DAF.below.cutoff  DAF.above.cutoff
#> -----
#> Neutral class          0          45101
#> Selected class         0          35816
#>
#> $`MKT tables`$`Number of segregating sites by DAF category - Cutoff = 0.05`
#>
#>
#> Table: cutoff
#>
#>      DAF.below.cutoff  DAF.above.cutoff
#> -----
#> Neutral class        17189        27912
#> Selected class       22490        13326
#>
#> $`MKT tables`$`Number of segregating sites by DAF category - Cutoff = 0.2`
#>
#>
#> Table: cutoff
#>
#>      DAF.below.cutoff  DAF.above.cutoff
#> -----
#> Neutral class        26931        18170
#> Selected class       28322         7494
#>
#> $`MKT tables`$`MKT standard table`
#>
#>
#>      Polymorphism  Divergence
#> -----
#> Neutral class    45101     52537
#> Selected class   35816     54641
#>
#>
#> $Fractions
#>      0      0.05      0.2
#> d 0.810538 0.81053943 0.81053627
#> f 0.189462 0.14269958 0.15277678
#> b 0.000000 0.04676099 0.03668695

```

Asymptotic MKT

Petrov reference + explanation

```

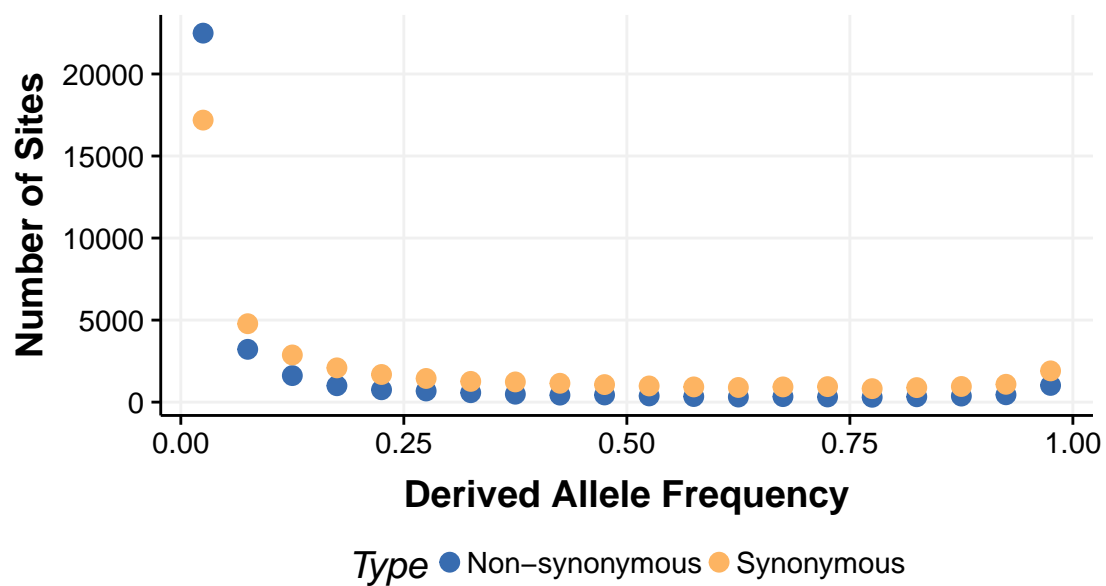
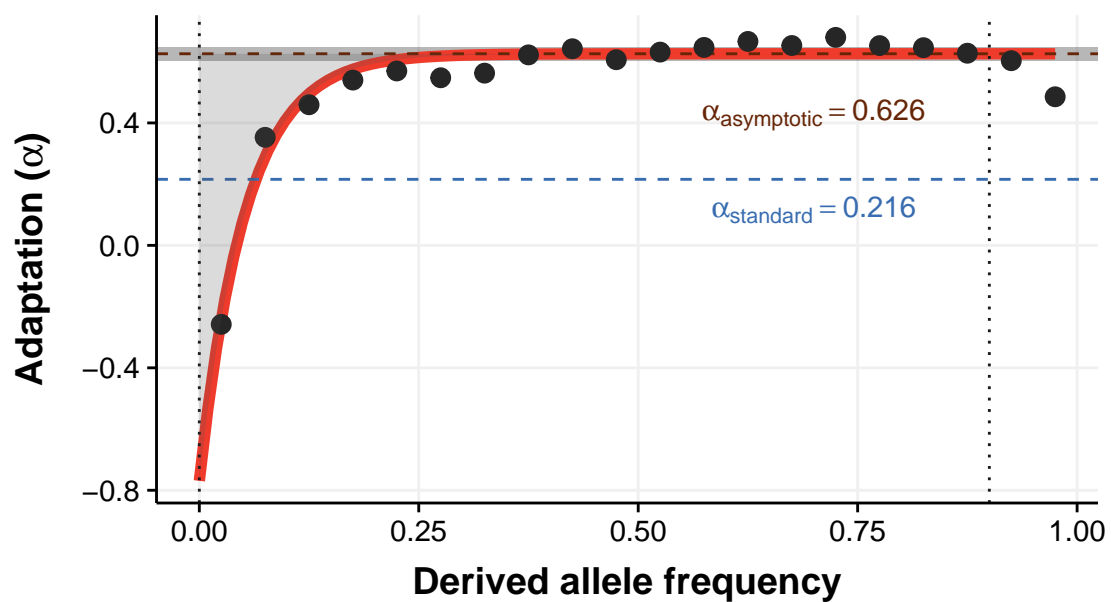
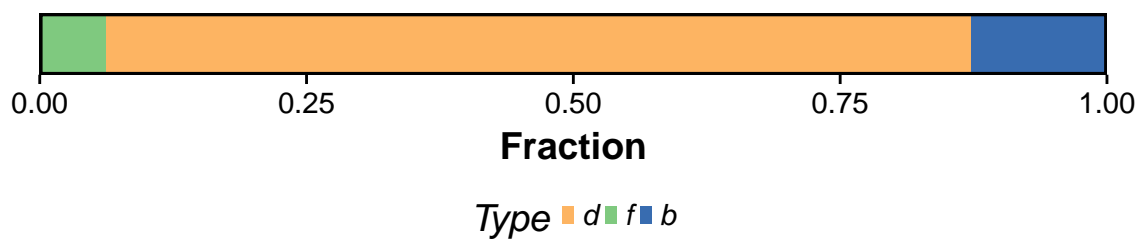
asymptoticMK(daf=mydafdata, divergence=mydivergencedata, xlow=0, xhigh=0.9)
#>      model      a      b      c alpha_asymptotic  CI_low
#> 1 exponential 0.6258904 -1.395108 18.96187      0.6258904 0.6039621
#>      CI_high alpha_original
#> 1 0.647417      0.2157308

```


iMK

Asymptotic explanation + Sergi slightly deleterious approach

```
iMK(daf=mydafdata, divergence=mydivergencedata, xlow=0, xhigh=0.9, plot=TRUE)
#> $`Asymptotic MK table`
#>      model      a      b      c alpha_asymptotic CI_low CI_high
#> 1 exponential 0.6259 -1.3951 18.9619      0.6259 0.6042 0.6467
#>      alpha_original
#> 1      0.2157
#>
#> $`Fractions of sites`
#>      Type      Fraction
#> 1      d 0.81053796
#> 2      f 0.06232362
#> 3      b 0.12713842
#>
#> $Graphs
```

A**B****C**

Miscelaneous

If you have a bunch of data like the following, or simply have several genes datasets: Maybe you want to perform some test or compare the test results between your datasets. You could execute the function `multipleDatasets`, putting your datasets in a directory and name them with the extensions **ID.daf.txt**/**ID.divergence.txt**. Then execute the following commands to perform the tests:

The `idList` argument allow to the user pass a plain text file with the IDs, in the case you want to subset the analysis to just a few datasets. It is used when `fullAnalysis = FALSE`, list of IDs to analyze