

Gamma Mixture Modelling

System Requirements:

MATLAB (2018b - v9.5)

Statistics and Machine Learning Toolbox (v11.4)

*Previous versions from 2016a onward have been tested

Installation:

Place the full package into a single directory, and open this directory in MATLAB.

Demo:

Load data (Msh2_Sample or WT_Sample) into a MATLAB variable named 'x'. The gamma mixture-model (MixModel.m) can then be ran (from the current directory) on the sample with the following syntax:

```
[alpha,beta,weight] = MixModel(x,2,500,0,'kmeans')
```

This will instantly return the variables:

alpha - estimates for both $\gamma(\alpha)$ values

beta - estimates for both $\gamma(\beta)$ values

weight - the contribution of each sub-distribution to the overall model (e.g. 0.45 = 45%)

Expected Results:

WT

alpha = [3.1437, 1.0991]

beta = [59118, 61354]

weight [0.6292, 0.3705]

i.e. class I = 62.92%, class II = 37.05%

Msh2

alpha = [3.6782, 0.9058]

beta = [31661, 42076]

weight [0.8480, 0.1518]

RecombineSim

System Requirements:

MATLAB (2018b - v9.5)

Statistics and Machine Learning Toolbox (v11.4)

Signal Processing Toolbox (v8.1)

*Previous versions from 2016a onward have been tested

Installation:

Place the full package into a single directory, and open this directory in MATLAB.

Demo:

Sample event table contains 4 WT (TW), 4 *ndt80AR* (TN) and 4 *msh2Δ* (OM) tetrads, at the 1500bp merging threshold. The following instructions are run in express mode (crossovers only) to reduce runtime.

Syntax:

```
RecombineSim(eventfile,varfile,output_folder,genotype,threshold,samples,mode,classIIperc,express,alpha,beta)
```

eventfile: SampleEventTable.txt

varfile: VariantTable_SK1.txt

output_folder: a directory or folder name

genotype: 'OM', 'TN' or 'TW'

threshold: 1500

samples: number of cells to simulate

mode: 'Random', 'Hazard' or 'MixModel' (see below)

classIIperc: the % of COs that will form independently of interference

express: 'Y'

alpha: the interfering $\gamma(\alpha)$ values for mixed model simulations

beta: the interfering $\gamma(\beta)$ values for mixed model simulations

To run a random simulation (1000 cells, Runtime = ~5-10 minutes):

```
RecombineSim('SampleEventTable.txt','VariantTable_SK1.txt','Results','OM',1500,1000,'Random',0,'Y')
```

'OM' can be replaced with 'TN' or 'TW' (as defined above) to simulate a different genotype.

To run an interfering, single (γ)-model simulation:

```
RecombineSim('SampleEventTable.txt','VariantTable_SK1.txt','Results','OM',1500,1000,'Hazard',0,'Y')
```

To run an interfering, mixed (γ)-model simulation:

```
RecombineSim('SampleEventTable.txt','VariantTable_SK1.txt','Results','OM',1500,1000,'MixModel',15,'Y',[3.5 3.5 3.5],[37500 37500 37500])
```

Where:

Class II % = 15

Interfering $\gamma(\alpha)$ = 3.5

Interfering $\gamma(\beta)$ = 37500

Output Folders:

Event Counts:

1. Event counts for each individual tetrad (e.g. OM-1500bp-EventCount.txt)

Experimental:

1. Experimental inter-event distances (IEDs) for each individual tetrad (e.g. OM-1500bp-ExpIED.txt)
2. Combined IED datasets (e.g. OM-1500bp-AggregateExpIED.txt) - the main dataset used in the manuscript
3. Best-fit (γ)-parameters for the experimental data (e.g. OM-1500bp-GammaFit.txt).

Simulations:

1. Simulated IED datasets for each individual tetrad (e.g. OM-Random-1500bp-PopAvg.txt)
2. Combined simulated IED datasets (e.g. OM-Random-1500bp-AggregateSim.txt) - the main dataset used in the manuscript
3. Duplication of the above files without event merging (Raw).