simuPOP Reference Manual

Release 0.8.8 (Rev: 1804)

Bo Peng

December 2004

Last modified
October 29, 2008

Department of Epidemiology, U.T. M.D. Anderson Cancer Center
Email: bpeng@mdanderson.org
URL: http://simupop.sourceforge.net
Mailing List: simupop-list@lists.sourceforge.net

Acknowledgements:

Dr. Marek Kimmel
Dr. François Balloux
Dr. William Amos
SWIG user community
Python user community
Keck Center for Computational and Structural Biology
U.T. M.D. Anderson Cancer Center

Abstract

simuPOP is a forward-time population genetics simulation environment. Unlike coalescent-based programs, simuPOP evolves populations forward in time, subject to arbitrary number of genetic and environmental forces such as mutation, recombination, migration and population/subpopulation size changes. Statistics of populations can be calculated and visualized dynamically which makes simuPOP an ideal tool to demonstrate population genetics models; generate datasets under various evolutionary settings, and more importantly, study complex evolutionary processes and evaluate gene mapping methods.

The core of simuPOP is a scripting language (Python) that provides a large number of building blocks (populations, mating schemes, various genetic forces in the form of operators, simulators and gene mapping methods) to construct a simulation. This provides a R/Splus or Matlab-like environment where users can interactively create, manipulate and evolve populations, monitor and visualize population statistics and apply gene mapping methods. The full power of simuPOP and Python (even R) can be utilized to simulate arbitrarily complex evolutionary scenarios.

simuPOP is written in C++ and is provided as Python modules. Besides a front-end providing an interactive shell and a scripting language, Python is used extensively to pass dynamic parameters, calculate complex statistics and write operators. Because of the openness of simuPOP and Python, users can make use of external programs, such as R, to perform statistical analysis, gene mapping and visualization. Depending on machine configuration, simuPOP can simulate large (think of millions) populations at reasonable speed.

This reference manual assumes that you are reasonably familiar with the Python programming language. If you are new to Python, you may want to go through a few online tutorials and courses before you continue. Because this is a reference manual to all the features of simuPOP, it is recommended that you learn the basic concepts of simuPOP from the *simuPOP User's Guide* before getting lost in the details.

This is a reference manual to all variables, functions, and objects of simuPOP. To learn different components of simuPOP and how to write simuPOP scripts, please refer to the *simuPOP User's Guide.*

How to cite simuPOP:

> Bo Peng and Marek Kimmel (2005) simuPOP: a forward-time population genetics simulation environment. *bioinformatics*, **21**(18): 3686-3687

CONTENTS

LIST OF EXAMPLES

CHAPTER

ONE

simuPOP Components

1.1 Individual and Population

1.1.1 Class `GenoStruTrait`

All individuals in a population share the same genotypic properties such as number of chromosomes, number and position of loci, names of alleles, markers, chromosomes, and information fields. These properties are stored in this `GenoStruTrait` class and are accessible from `individual`, `population`, and `simulator` classes. Currently, a genotypic structure consists of

- Ploidy, namely the number of homologous sets of chromosomes, of a population. Haplodiploid population is also supported.
- Number of chromosomes and number of loci on each chromosome.
- Positions of loci, which determine the relative distance between loci on the same chromosome. No unit is assumed so these positions can be ordinal (`1`, `2`, `3`, ..., the default), in physical distance (`bp`, `kb` or `mb`), or in map distance (e.g. centiMorgan) depending on applications.
- Names of alleles. Although alleles at different loci usually have different names, simuPOP uses the same names for alleles across loci for simplicity.
- Names of loci and chromosomes.
- Names of information fields attached to each individual.

In addition to basic property access functions, this class also provides some utility functions such as `locusByName`, which looks up a locus by its name.

class `GenoStruTrait`()

A `GenoStruTrait` object is created with the creation of a `population` so it cannot be initialized directly.

`alleleName`(*allele*)

Return the name of allele *allele* specified by the *alleleNames* parameter of the `population` function. If the name of an allele is not specified, its index (`'0'`, `'1'`, `'2'`, etc) is returned.

`alleleNames`()

Return a list of allele names given by the *alleleNames* parameter of the `population` function. This list does not have to cover all possible allele states of a population.

`maxAllele`()

Return the maximum allowed allele state of the current simuPOP module, which is `1` for binary modules, `255` for short modules and `65535` for long modules.

chromBegin(*chrom*)
Return the index of the first locus on chromosome *chrom*.

chromByName(*name*)
Return the index of a chromosome by its *name*.

chromEnd(*chrom*)
Return the index of the last locus on chromosome *chrom* plus 1.

chromName(*chrom*)
Return the name of a chromosome *chrom*. Default to `chrom#` where # is the 1-based index of the chromosome.

chromNames()
Return a tuple of the names of all chromosomes.

hasSexChrom()
Return `True` if the last chromosome is the sex chromosome.

numChrom()
Return the number of chromosomes.

infoField(*idx*)
Return the name of information field *idx*.

infoFields()
Return a tuple of the names of all information fields of the population.

infoIdx(*name*)
Return the index of information field *name*. Raise an `IndexError` if *name* is not one of the information fields.

absLocusIndex(*chrom, locus*)
Return the absolute index of locus *locus* on chromosome *chrom*. An `IndexError` will be raised if *chrom* or *locus* is out of range. c.f. `chromLocusPair`.

chromLocusPair(*locus*)
Return the chromosome and relative index of a locus using its absolute index *locus*. c.f. `absLocusIndex` .

lociByNames(*names*)
Return the indexes of loci with names *names*. Raise a `ValueError` if any of the loci cannot be found.

lociDist(*loc1, loc2*)
Return the distance between loci *loc1* and *loc2* on the same chromosome. A negative value will be returned if *loc1* is after *loc2*.

lociNames()
Return the names of all loci specified by the *lociNames* parameter of the `population` function.

lociPos()
Return the positions of all loci, specified by the *lociPos* prameter of the `population` function. The default positions are 1, 2, 3, 4, ... on each chromosome.

locusByName(*name*)
Return the index of a locus with name *name*. Raise a `ValueError` if no locus is found.

locusName(*loc*)
Return the name of locus *loc* specified by the *lociNames* parameter of the `population` function. Default to `locX-Y` where `X` and `Y` are 1-based chromosome and locus indexes (`loc1-1`, `loc1-2`, ... etc)

locusPos(*loc*)
Return the position of locus *loc* specified by the *lociPos* parameter of the `population` function. An `IndexError` will be raised if the absolute index *loc* is greater than or equal to the total number of loci.

numLoci(*chrom*)
Return the number of loci on chromosome *chrom*, equivalent to `numLoci()[`*chrom*`]`.

numLoci()
Return the number of loci on all chromosomes.

totNumLoci()
Return the total number of loci on all chromosomes.

isHaplodiploid()
Return `True` if this population is haplodiploid.

ploidy()
Return the number of homologous sets of chromosomes, specified by the *ploidy* parameter of the `population` function. Return 2 for a haplodiploid population because two sets of chromosomes are stored for both males and females in such a population.

ploidyName()
Return the ploidy name of this population, can be one of `haploid`, `diploid`, `haplodiploid`, `triploid`, `tetraploid` or `#-ploid` where # is the ploidy number.

1.1.2 Class `individual`

Individuals with genotype, affection status, sex etc. Individuals are the building blocks of populations, each having the following individual information:

- shared genotypic structure information
- genotype
- sex, affection status, subpopulation ID
- optional information fields

Individual genotypes are arranged by locus, chromosome, ploidy, in that order, and can be accessed from a single index. For example, for a diploid individual with two loci on the first chromosome, one locus on the second, its genotype is arranged as `1-1-1 1-1-2 1-2-1 2-1-1 2-1-2 2-2-1` where `x-y-z` represents ploidy `x` chromosome `y` and locus `z`. An allele `2-1-2` can be accessed by `allele(4)` (by absolute index), `allele(1, 1)` (by index and ploidy) or `allele(1, 1, 0)` (by index, ploidy and chromosome). Individuals are created by populations automatically. Do not call the constructor function directly.

class individual()
FIXME: No document

affected()
Whether or not an individual is affected

affectedChar()
Return `A` (affected) or `U` (unaffected) for affection status

allele(*index*)
Return the allele at locus `index`

index: Absolute index from the beginning of the genotype, ranging from `0` to `totNumLoci()*ploidy()`

allele(*index, p*)
Return the allele at locus `index` of the `p-th` copy of the chromosomes

index: Index from the begining of the `p-th` set of the chromosomes, ranging from `0` to `totNumLoci()`

p: Index of the ploidy

allele(*index, p, ch*)
Return the allele at locus `index` of the `ch-th` chromosome in the `p-th` chromosome set
ch: Index of the chromosome in the `p-th` chromosome set
index: Index from the begining of chromosome `ch` of ploidy `p`, ranging from `0` to `numLoci(ch)`
p: Index of the polidy

alleleChar(*index*)
Return the name of `allele(index)`

alleleChar(*index, p*)
Return the name of `allele(index, p)`

alleleChar(*index, p, ch*)
Return the name of `allele(idx, p, ch)`

genotype()
Return an editable array (a carray of length `totNumLoci()*ploidy()`) of genotypes of an individual.

genotype(*p*)
Return an editable array of alleles of the `p-th` copy of the chromosomes
p: Index of the ploidy

genotype(*p, ch*)
Return an editable array of alleles of the `ch-th` chromosome in the `p-th` chromosome set
ch: Index of the chromosome in ploidy `p`
p: Index of the ploidy

info(*idx*)
Get information field `idx`
idx: Index of the information field

info(*name*)
Get information field `name`Equivalent to `info(infoIdx(name))`.
name: Name of the information field

intInfo(*idx*)
Get information field `idx` as an integer. This is the same as `int(info(idx))`
idx: Index of the information field

intInfo(*name*)
Get information field `name` as an integer Equivalent to `int(info(name))`.
name: Name of the information field

setAffected(*affected*)
Set affection status

setAllele(*allele, index*)
Set the allele at locus `index`
allele: Allele to be set
index: Index from the begining of genotype, ranging from `0` to `totNumLoci()*ploidy()`

setAllele(*allele, index, p*)
Set the allele at locus `index` of the `p-th` copy of the chromosomes
allele: Allele to be set
index: Index from the begining of the ploidy `p`, ranging from `0` to `totNumLoci(p)`
p: Index of the ploidy

setAllele(*allele, index, p, ch*)
Set the allele at locus `index` of the `ch-th` chromosome in the `p-th` chromosome set
allele: Allele to be set
ch: Index of the chromosome in ploidy `p`

index: Index from the begining of the chromosome, ranging from `0` to `numLoci(ch)`

p: Index of the ploidy

setGenotype(*geno*)

Set the genotype of an individual

geno: Genotype to be set. It will be reused if its length is less than the genotype length of the individual.

setGenotype(*geno, p*)

Set the genotype of the `p-th` copy of the chromosomes

geno: Genotype to be set. It will be reused if its length is less than the total number of loci.

p: Index of the ploidy

setGenotype(*geno, p, ch*)

Set the genotype of the `ch-th` chromosome in the `p-th` chromosome set

ch: Index of the chromosome in ploidy `p`

geno: Genotype to be set. It will be reused if its length is less than the number of loci on chromosome `ch`.

p: Index of the ploidy

setInfo(*value, idx*)

Set information field by `idx`

setInfo(*value, name*)

Set information field by `name`

setSex(*sex*)

Set sex. `sex` can be `Male` of `Female`.

setSubPopID(*id*)

Set new subpopulation ID, `pop.rearrangeByIndID` will move this individual to that population

sex()

Return the sex of an individual, `1` for males and `2` for females.

sexChar()

Return the sex of an individual, `M` or `F`

subPopID()

Return the ID of the subpopulation to which this individual blongs

Note: `subPopID` is not set by default. It only corresponds to the subpopulation in which this individual resides after `pop::setIndSubPopID` is called.

unaffected()

Equals to `not affected()`

1.1.3 Class `population`

A collection of individuals with the same genotypic structure. A simuPOP population consists of individuals of the same genotypic structure, which refers to the number of chromosomes, numbers and positions of loci on each chromosome etc. The most important components of a population are:

- subpopulations. A population is divided into subpopulations (unstructured population has a single subpopulation, which is the whole population itself). Subpopulation structure limits the usually random exchange of genotypes between individuals by disallowing mating between individuals from different subpopulations. In the presence of subpopualtion structure, exchange of genetic information across subpopulations can only be done through migration. Note that simuPOP uses one-level population structure, which means there is no sub-subpopulation or family in subpopulations.

- variables. Every population has its own variable space, or *local namespace* in simuPOP term. This namespace is a Python dictionary that is attached to each population and can be exposed to the users through

`vars()` or `dvars()` function. Many functions and operators work and store their results in this namespace. For example, function `Stat` sets variables such as `alleleFreq[loc]`, and you can access it via `pop.dvars().alleleFreq[loc][allele]`.

- ancestral generations. A population can save arbitrary number of ancestral generations. During evolution, the latest several (or all) ancestral generations are saved. Functions to switch between ancestral generations are provided so that one can examine and modify ancestral generations.

class `population` (*size=[], ploidy=2, loci=[], sexChrom=False, lociPos=[], ancestralDepth=0, chromNames=[], alleleNames=[], lociNames=[], maxAllele=ModuleMaxAllele, infoFields=[]*)

alleleNames: An array of allele names. For example, for a locus with alleles A, C, T, G, you can specify `alleleNames` as `('A','C','T','G')`.

ancestralDepth: Number of most recent ancestral generations to keep during evolution. Default to `0`, which means only the current generation will be available. You can set it to a positive number `m` to keep the latest m generations in the population, or `-1` to keep all ancestral populations. Note that keeping track of all ancestral generations may quickly exhaust your computer RAM. If you really need to do that, using `savePopulation` operator to save each generation to a file is a much better choice.

chromNames: An array of chromosome names.

infoFields: Names of information fields that will be attached to each individual. For example, if you need to record the parents of each individual using operator `parentTagger()`, you will need two fields `father_idx` and `mother_idx`.

loci: An array of numbers of loci on each chromosome. The length of parameter `loci` determines the number of chromosomes. Default to `[1]`, meaning one chromosome with a single locus.

The last chromosome can be sex chromosome. In this case, the maximum number of loci on X and Y should be provided. I.e., if there are 3 loci on Y chromosme and 5 on X chromosome, use `5`.

lociNames: An array or a matrix (separated by chromosomes) of names for each locus. Default to `"locX-Y"` where `X` is the chromosome index and `Y` is the locus number, both starting from 1.

lociPos: A 1-d or 2-d array specifying positions of loci on each chromosome. You can use a nested array to specify loci position for each chromosome. For example, you can use `lociPos=[1,2,3]` when `loci=[3]` or `lociPos=[[1,2],[1.5,3,5]]` for `loci=[2,3]`. simuPOP does not assume a unit for these positions, although they are usually intepreted as centiMorgans. The default values are `1`, `2`, etc. on each chromosome.

maxAllele: Maximum allele number. Default to the maximum allowed allele state of the current library. This will set a cap for all loci. For individual locus, you can specify `maxAllele` in mutation models, which can be smaller than the global `maxAllele` but not larger. Note that this number is the number of allele states minus 1 since allele number starts from 0.

ploidy: Number of sets of homologous copies of chromosomes. Default to `2` (diploid). Please use `Haplodiploid` to specify a haplodiploid population. Note that the ploidy number returned for such a population will be 2 and male individuals will store two copies of chromosomes. Operators such as a recombinator will recognize this population as haplodiploid and act accordingly.

sexChrom: Diploid population only. If this parameter is `True`, the last homologous chromosome will be treated as sex chromosome. (XY for male and XX for female.) If X and Y have different numbers of loci, the number of loci of the longer one of the last (sex) chromosome should be specified in `loci`.

size: An array of subpopulation sizes. If a single number is given, it will be the size of a single subpopulation of the whole population.

subPop: Obsolete parameter

Create a population object with given size and genotypic structure.

absIndIndex(*ind, subPop*)

Return the absolute index of an individual in a subpopulation.

index: Index of an individual in a subpopulation `subPop`

subPop: Subpopulation index (start from `0`)

addInfoField(*field, init=0*)

Add an information field to a population

field: New information field. If it already exists, it will be re-initialized.

init: Initial value for the new field.

addInfoFields(*fields, init=0*)

Add one or more information fields to a population

fields: An array of new information fields. If one or more of the fields alreay exist, they will be re-initialized.

init: Initial value for the new fields.

ancestor(*ind, gen*)

Refrence to an individual `ind` in an ancestral generation This function gives access to individuals in an ancestral generation. It will refer to the correct generation even if the current generation is not the latest one. That is to say, ancestor(ind, 0) is not always individual(ind).

ancestor(*ind, subPop, gen*)

Refrence to an individual `ind` in a specified subpopulaton or an ancestral generation This function gives access to individuals in an ancestral generation. It will refer to the correct generation even if the current generation is not the latest one. That is to say, ancestor(ind, 0) is not always individual(ind).

ancestralDepth()

Ancestral depth of the current population

Note: The return value is the number of ancestral generations exist in the population, not necessarily equals to the number set by `setAncestralDepth()`.

ancestralGen()

Currently used ancestral population (`0` for the latest generation) Current ancestral population activated by `useAncestralPop()`. There can be several ancestral generations in a population. `0` (current), `1` (parental) etc. When `useAncestralPop(gen)` is used, current generation is set to one of the parental generations, which is the information returned by this function. `useAncestralPop(0)` should always be used to set a population to its usual ancestral order after operations to the ancestral generation are done.

clone(*keepAncestralPops=-1*)

Deep copy of a population. (In python, `pop1 = pop` will only create a reference to `pop`.) This function by default copies all ancestral generations, but you can copy only one (current, `keepAncestralPops=0`), or specified number of ancestral generations.

deactivateVirtualSubPop(*subPop*)

Deactivate virtual subpopulations in a given subpopulation. In another word, all individuals will become visible.

Note: this function is currently not recommended to be used.

evaluate(*expr="", stmts=""*)

Evaluate a Python statment/expression in the population's local namespace This function evaluates a Python statment(`stmts`)/expression(`expr`) and return its result as a string. Optionally run statement(`stmts`) first.

execute(*stmts=""*)

Execute a statement (can be a multi-line string) in the population's local namespace

fitSubPopStru(*newSubPopSizes*)

Of the population will be cleared.

gen()

Current generation during evolution

genotype()
Get an editable array of the genotype of all individuals in a population. Return an editable array of all genotypes of the population. You need to know how these genotypes are organized to safely read/write genotype directly.

genotype(*subPop*)
Get an editable array of the genotype of all individuals in a subpopulation.

subPop: Index of subpopulation (start from 0)

hasVirtualSubPop()
If a population has any virtual subpopulation

indBegin(*subPop*)
The iterator will skip invisible individuals.

indEnd()
It is recommended to use it.valid(), instead of it != indEnd().

indEnd(*subPop*)
It is recommended to use it.valid(), instead of it != indEnd(sp).

indInfo(*idx*)
Get information field `idx` of all individuals

idx: Index of the information field

indInfo(*name*)
Get information field `name` of all individuals

name: Name of the information field

indInfo(*idx, subPop*)
Get information field `idx` of all individuals in a subpopulation `subPop`

idx: Index of the information field

subPop: Subpopulation index

indInfo(*name, subPop*)
Get information field `name` of all individuals in a subpopulation `subPop`

name: Name of the information field

subPop: Subpopulation index

individual(*ind, subPop=0*)
Refernce to individual `ind` in subpopulation `subPop`This function is named `individual` in the Python interface.

ind: Individual index within `subPop`

subPop: Subpopulation index

individuals()
Return an iterator that can be used to iterate through all individuals Typical usage is

```
for ind in pop.individuals():
```

individuals(*subPop*)
Return an iterator that can be used to iterate through all individuals in subpopulation `subPop`

individuals(*subPop, virtualSubPop*)
FIXME: No document

insertAfterLoci(*idx, pos, names=[]*)
Append loci after given positions Append loci at some given locations. Alleles at inserted loci are initialized with zero allele.

idx: An array of locus index. The loci will be added *after* each index. If you need to append to the first locus of a chromosome, use `insertBeforeLoci` instead. If your index is the last locus of a chromosome, the appended locus will become the last locus of that chromosome. If you need to append multiple loci after a locus, repeat that locus number.

names: An array of locus names. If this parameter is not given, some unique names such as "insX_Y" will be given.

pos: An array of locus positions. The positions of the appended loci have to be between adjacent markers.

insertAfterLocus (*idx, pos, name=string*)

Append an locus after a given position `insertAfterLocus(idx, pos, name)` is a shortcut to `insertAfterLoci([idx], [pos], [name])`.

insertBeforeLoci (*idx, pos, names=[]*)

Insert loci before given positions Insert loci at some given locations. Alleles at inserted loci are initialized with zero allele.

idx: An array of locus index. The loci will be inserted *before* each index. If you need to append to the last locus, use `insertAfterLoci` instead. If your index is the first locus of a chromosome, the inserted locus will become the first locus of that chromosome. If you need to insert multiple loci before a locus, repeat that locus number.

names: An array of locus names. If this parameter is not given, some unique names such as "insX_Y" will be given.

pos: An array of locus positions. The positions of the appended loci have to be between adjacent markers.

insertBeforeLocus (*idx, pos, name=string*)

Insert an locus before a given position `insertBeforeLocus(idx, pos, name)` is a shortcut to `insertBeforeLoci([idx], [pos], [name])`

locateRelatives (*relType, relFields, gen=-1, relSex=AnySex, parentFields=[]*)

Find relatives of each individual and fill the given information fields with their indexes. This function locates relatives of each individual and store their indexes in given information fields.

gen: Find relatives for individuals for how many generations. Default to -1, meaning for all generations. If a non-negative number is given, up till generation gen will be processed.

parentFields: Information fields that stores parental indexes. Default to ['father_idx', 'mother_idx']

relFields: Information fields to hold relatives. The number of these fields limits the number of relatives to locate.

relSex: Whether or not only locate relative or certain sex. It can be AnySex (do not care, default), MaleOnly, FemaleOnly, or OppositeSex (only locate relatives of opposite sex.

relType: Relative type, can be

- REL_Self index of individual themselfs
- REL_Spouse index of spouse in the current generation. Spouse is defined as two individuals having an offspring with shared `parentFields`. If more than one `infoFields` is given, multiple spouses can be identified.
- REL_Offspring index of offspring in the offspring generation. If only one parent is given, only paternal or maternal relationship is considered. For example, `parentFields=['father_idx']` will locate offspring for all fathers.
- REL_FullSibling all siblings with the same parents
- REL_Sibling all sibs with at least one shared parent

mergePopulation (*pop, newSubPopSizes=[], keepAncestralPops=-1*)

Merge populations by individuals Merge individuals from `pop` to the current population. Two populations should have the same genotypic structures. By default, subpopulations of the merged populations are kept. I.e., if you merge two populations with one subpopulation, the resulting population will have two subpopulations. All ancestral generations are also merged by default.

keepAncestralPops: Ancestral populations to merge, default to all (`-1`)

newSubPopSizes: Subpopulation sizes can be specified. The overall size should be the combined size of the two populations. Because this parameter will be used for all ancestral generations, it may fail if ancestral generations have different sizes. To avoid this problem, you can run `mergePopulation` without this parameter, and then adjust subpopulation sizes generation by generation.

Note: Population variables are not copied to `pop`.

mergePopulationByLoci(*pop, newNumLoci=[], newLociPos=[], byChromosome=False*)

Merge populations by loci Two populations should have the same number of individuals. This also holds for any ancestral generations. By default, chromosomes of `pop` are appended to the current population. You can change this arrangement in two ways

- specify new chromosome structure using parameter `newLoci` and `newLociPos`. Loci from new and old populations are still in their original order, but chromosome number and positions can be changed in this way.
- specify `byChromosome=true` so that chromosomes will be merged one by one. In this case, loci position of two popualtions are important because loci will be arranged in the order of loci position; and identical loci position of two loci in two populations will lead to error.

byChromosome: Merge chromosome by chromosome, loci are ordered by loci position Default to `False`.

newLociPos: The new loci position if number of loci on each chromosomes are changed with `newNumLoci`. New loci positions should be in order on the new chromosomes.

newNumLoci: The new number of loci for the combined genotypic structure.

Note:

- Information fields are not merged.
- All ancestral generations are merged because all individuals in a population have to have the same genotypic structure.

mergeSubPops(*subPops=[], removeEmptySubPops=False*)

Merge given subpopulations Merge subpopulations, the first subpopulation ID (the first one in array `subPops`) will be used as the ID of the new subpopulation. That is to say, all merged subpopulations will take the ID of the first one. The subpopulation ID of the empty subpopulations will be kept (so that other subpopulations are unaffected, unless they are removed by `removeEmptySubPops = True`).

newPopByIndID(*keepAncestralPops=-1, id=[], removeEmptySubPops=False*)

Form a new population according to individual subpopulation ID. Individuals with negative subpopulation ID will be removed.

newPopWithPartialLoci(*remove=[], keep=[]*)

Obtain a new population with selected loci Copy current population to a new one with selected loci `keep` or remove specified loci `remove` (no change on the current population), equivalent to

```
y=x.clone
y.removeLoci(remove, keep)
```

numSubPop()

Number of subpopulations in a population.

numVirtualSubPop()

Number of virtual subpopulations.

popSize()

Total population size

pushAndDiscard(*rhs, force=False*)

Absorb `rhs` population as the current generation of a population This function is used by a simulator to push offspring generation `rhs` to the current population, while the current population is pushed back as an ancestral population (if `ancestralDepath() != 0`). Because `rhs` population is swapped in, `rhs` will be empty after this operation.

rawIndBegin(*subPop*)

The iterator will skip invisible individuals.

rawIndEnd()

It is recommended to use it.valid(), instead of it != indEnd().

rawIndEnd(*subPop*)

It is recommended to use it.valid(), instead of it != indEnd(sp).

removeEmptySubPops()
Remove empty subpopulations by adjusting subpopulation IDs

removeIndividuals(*inds=[], subPop=-1, removeEmptySubPops=False*)
Remove individuals. If a valid `subPop` is given, remove individuals from this subpopulation. Indexes in `inds` will be treated as relative indexes.

removeLoci(*remove=[], keep=[]*)
Remove some loci from the current population. Only one of the two parameters can be specified.

removeSubPops(*subPops=[], shiftSubPopID=True, removeEmptySubPops=False*)
Remove subpopulations and adjust subpopulation IDs so that there will be no *'empty'* subpopulation left Remove specified subpopulations (and all individuals within). If `shiftSubPopID` is false, `subPopID` will be kept intactly.

reorderSubPops(*order=[], rank=[], removeEmptySubPops=False*)
Reorder subpopulations by `order` or by `rank`

order: New order of the subpopulations. For examples, 3 2 0 1 means `subpop3`, `subpop2`, `subpop0`, `subpop1` will be the new layout.

rank: You may also specify a new rank for each subpopulation. For example, 3,2,0,1 means the original subpopulations will have new IDs 3,2,0,1, respectively. To achive order 3,2,0,1, the rank should be 1 0 2 3.

rep()
Current replicate in a simulator which is not meaningful for a stand-alone population

resize(*newSubPopSizes, propagate=False*)
Resize current population Resize population by giving new subpopulation sizes.

newSubPopSizes: An array of new subpopulation sizes. If there is only one subpopulation, use `[newPopSize]`.

propagate: If `propagate` is `true`, copy individuals to new comers. I.e., 1, 2, 3 ==> 1, 2, 3, 1, 2, 3, 1

Note: This function only resizes the current generation.

savePopulation(*filename, format="", compress=True*)
Save population to a file

compress: Obsolete parameter

filename: Save to filename

format: Obsolete parameter

selectionOn()
Selection is on at any subpopulation?

setAncestralDepth(*depth*)
Set ancestral depth

depth: `0` for none, `-1` for unlimited, a positive number sets the number of ancestral generations to save.

setGenotype(*geno*)
Set genotype to all individuals of a population.

geno: Genotype to be set. It will be reused if its length is less than the genotype length of the population, which is popSize()*ploidy()*totNumLoci().

setGenotype(*geno, subPop*)
Set genotype to all individuals in a subpopulation.

geno: Genotype to be set. It will be reused if its length is less than the genotype length of the population, which is subPopSize(subPop)*ploidy()*totNumLoci().

subPop: Index of subpopulation (start from 0)

setIndInfo(*values, idx*)
Set individual information for the given information field `idx`

idx: Index to the information field.

values: An array that has the same length as population size.

setIndInfo(*values, name*)

Set individual information for the given information field `namex.setIndInfo(values, name)` is equivalent to the `idx` version `x.setIndInfo(values, x.infoIdx(name))`.

setIndSubPopID(*id, ancestralPops=False*)

Set subpopulation ID with given ID Set subpopulation ID of each individual with given ID. Individuals can be rearranged afterwards using `setSubPopByIndID`.

ancestralPops: If true (default to False), set subpop id for ancestral generations as well.

id: An array of the same length of population size, resprenting subpopulation ID of each individual. If the length of is less than population size, it is repeated to fill the whole population.

setIndSubPopIDWithID(*ancestralPops=False*)

Set subpopulation ID of each individual with their current subpopulation ID

ancestralPops: If true (default to False), set subpop id for ancestral generations as well.

setIndexesOfRelatives(*pathGen, pathFields, pathSex=[], resultFields=[]*)

Trace a relative path in a population and record the result in the given information fields. For example, `setInfoWithRelatives(pathGen = [0, 1, 1, 0], pathFields = [['father_idx', 'mother_idx'], ['sib1', 'sib2'], ['off1', 'off2']], pathSex = [AnySex, MaleOnly, FemaleOnly], resultFields = ['cousin1', 'cousin2'])` This function will 1. locate father_idx and mother_idx for each individual at generation 0 (`pathGen`[0]) 2. find AnySex individuals referred by father_idx and mother_idx at generation 1 (`pathGen`[1]) 3. find informaton fields `sib1` and `sib2` from these parents 4. locate MaleOnly individuals referred by `sib1` and `sib2` from generation 1 (`pathGen`[2]) 5. find information fields `off1` and `off2` from these individuals, and 6. locate FemaleOnly indiviudals referred by `off1` and from geneartion 0 (`pathGen`[3]) 7. Save index of these individuals to information fields `cousin1` and `cousin2` at genearation `pathGen`[0].In short, this function locates father or mother's brother's daughters.

pathFields: A list of list of information fields forming a path to trace a certain type of relative.

pathGen: A list of generations that form a relative path. This array is one element longer than `pathFields`, with gen_i, gen_i+1 indicating the current and destinating generation of information fields path_i.

pathSex: (Optional) A list of sex choices, AnySex, Male, Female or OppositeSex, that is used to choose individuals at each step. Default to AnySex.

resultFields: Where to store located relatives. Note that the result will be saved in the starting generation specified in `pathGen`[0], which is usually 0.

setInfoFields(*fields, init=0*)

Set information fields for an existing population. The existing fields will be removed.

fields: An array of fields

init: Initial value for the new fields.

setSubPopByIndID(*id=[]*)

Move individuals to subpopulations according to individual subpopulation IDs Rearrange individuals to their new subpopulations according to their subpopulation ID (or the new given `ID`). Order within each subpopulation is not respected.

id: New subpopulation ID, if given, current individual subpopulation ID will be ignored.

Note: Individual with negative info will be removed!

setSubPopStru(*newSubPopSizes*)

Set population/subpopulation structure given subpopulation sizes

newSubPopSizes: An array of new subpopulation sizes. The overall population size should not changed.

setVirtualSplitter(*vsp*)

FIXME: No document

vsp: A virtual subpop splitter

splitSubPop(*which, sizes, subPopID=[]*)
Split a subpopulation into subpopulations of given sizes The sum of given sizes should be equal to the size of the split subpopulation. Subpopulation IDs can be specified. The subpopulation IDs of non-split subpopulations will be kept. For example, if subpopulation 1 of 0 1 2 3 is split into three parts, the new subpop id will be 0 (1 4 5) 2 3.

Note: `subpop` with negative ID will be removed. So, you can shrink one `subpop` by splitting and setting one of the new `subpop` with negative ID.

splitSubPopByProportion(*which, proportions, subPopID=[]*)
Split a subpopulation into subpopulations of given proportions The sum of given proportions should add up to one. Subpopulation IDs can be specified.

Note: `subpop` with negative ID will be removed. So, you can shrink one `subpop` by splitting and setting one of the new `subpop` with negative ID.

subPopBegin(*subPop*)
Index of the first individual of a subpopulation `subPop`

subPopEnd(*subPop*)
Return the value of the index of the last individual of a subpopulation `subPop` plus 1

subPopIndPair(*ind*)
Return the subpopulation ID and relative index of an individual with absolute index `ind`

subPopSize(*subPop*)
Return size of a subpopulation `subPop`.

subPop: Index of subpopulation (start from 0)

subPopSizes()
Return an array of all subpopulation sizes.

swap(*rhs*)
Swap the content of two populations

turnOffSelection()
Turn off selection for all subpopulations This is only used when you would like to apply two selectors. Maybe using two different information fields.

turnOnSelection()
Turn on selection for all subpopulations.

useAncestralPop(*idx*)
Use an ancestral generation. `0` for the latest generation.

idx: Index of the ancestral generation. `0` for current, `1` for parental, etc. idx can not exceed ancestral depth (see setAncestralDepth).

validate(*msg*)
Evolution

vars(*subPop=-1*)
Return variables of a population. If `subPop` is given, return a dictionary for specified subpopulation.

virtualSubPopName(*subPop, virtualSubPop=InvalidSubPopID*)
Name of the given virtual subpopulation.

id: Subpopulation id

vid: Virtual subpopulation id

virtualSubPopSize(*subPop, virtualSubPop=InvalidSubPopID*)
Return the size of virtual subpopulation subPop. if subPop is activated, and subPop does not specify which virtual subpopulation to count, the currently activated virtual subpop is returned. Therefore, When it is not certain if a subpopulation has activated virtual subpopulation, this function can be used.

id: Subpopulation id

vid: Virtual subpopulation id. If not given, current subpopulation, or current actived subpopulation size will be returned.

1.1.4 Class `sexSplitter`

Split the population into Male and Female virtual subpopulations

class sexSplitter()
FIXME: No document

clone()
FIXME: No document

name(*sp*)
Name of a virtual subpopulation

numVirtualSubPop()
Number of virtual subpops of subpopulation sp

1.1.5 Class `affectionSplitter`

Split a subpopulation into unaffected and affected virtual subpopulations.

class affectionSplitter()
FIXME: No document

clone()
FIXME: No document

name(*sp*)
Name of a virtual subpopulation

numVirtualSubPop()
Number of virtual subpops of subpopulation sp

1.1.6 Class `infoSplitter`

Split the population according to the value of an information field. A list of distinct values, or a cutoff vector can be given to determine how the virtual subpopulations are divided. Note that in the first case, an individual does not have to belong to any virtual subpopulation.

class infoSplitter(*info, values=[]nfo, cutoff=[]*)
cutoff: A list of cutoff values. For example, cutoff=[1, 2] defines three virtual subpopulations with v < 1, 1 <= v < 2, and v >= 2.

info: Name of the information field

values: A list of values, each defines a virtual subpopulation

FIXME: No document

name(*sp*)
Name of a virtual subpopulation

numVirtualSubPop()
Number of virtual subpops of subpopulation sp

1.1.7 Class `proportionSplitter`

Split the population according to a proportion

class proportionSplitter(*proportions=[]*)

proportions: A list of float numbers (between 0 and 1) that defines the proportion of individuals in each virtual subpopulation. These numbers should add up to one.

FIXME: No document

clone()

FIXME: No document

name(*sp*)

Name of a virtual subpopulation

numVirtualSubPop()

Number of virtual subpops of subpopulation sp

1.1.8 Class rangeSplitter

Split the population according to individual range. The ranges can overlap and does not have to add up to the whole subpopulation.

class rangeSplitter(*ranges*)

range: A shortcut for ranges=[range]

ranges: A list of ranges

FIXME: No document

clone()

FIXME: No document

name(*sp*)

Name of a virtual subpopulation

numVirtualSubPop()

Number of virtual subpops of subpopulation sp

1.1.9 Class genotypeSplitter

Split the population according to given genotype

class genotypeSplitter(*loci, alleles, phase=False*)

alleles: A list (for each virtual subpopulation), of a list of alleles at each locus. If phase if true, the order of alleles is significant. If more than one set of alleles are given, individuals having either of them is qualified.

loci: A list of locus at which alleles are used to classify individuals

locus: A shortcut to loci=[locus]

phase: Whether or not phase is respected.

For example, Genotype Aa or aa at locus 1: locus = 1, alleles = [0, 1] Genotype Aa at locus 1 (assuming A is 1): locus = 1, alleles = [1, 0], phase = True Genotype AaBb at loci 1 and 2: loci = [1, 2], alleles = [1, 0, 1, 0], phase = True Two virtual subpopulations with Aa and aa locus = 1, alleles = [[1, 0], [0, 0]], phase = True A virtual subpopulation with Aa or aa locus = 1, alleles = [1, 0, 0, 0] Two virtual subpopulation with genotype AA and the rest locus = 1, alleles = [[1, 1], [1, 0, 0, 0]], phase = False

clone()

FIXME: No document

name(*sp*)

Name of a virtual subpopulation

numVirtualSubPop()
Number of virtual subpops of subpopulation sp

1.1.10 Class `combinedSplitter`

This plitter takes several splitters, and stacks their virtual subpopulations together. For example, if the first splitter has three vsp, the second has two. The two vsp from the second splitter will be the fouth (index 3) and fifth (index 4) of the combined splitter.

class combinedSplitter(*splitters=[]*)
FIXME: No document

clone()
FIXME: No document

name(*sp*)
Name of a virtual subpopulation

numVirtualSubPop()
Number of virtual subpops of subpopulation sp

1.2 Mating Scheme

1.2.1 Class `mating`

The base class of all mating schemes - a required parameter of `simulator`Mating schemes specify how to generate offspring from the current population. It must be provided when a simulator is created. Mating can perform the following tasks:

- change population/subpopulation sizes;
- randomly select parent(s) to generate offspring to populate the offspring generation;
- apply *during-mating* operators;
- apply selection if applicable.

class mating(*newSubPopSize=[], newSubPopSizeExpr="", newSubPopSizeFunc=None, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)
newSubPopSize: An array of subpopulations sizes, should have the same number of subpopulations as the current population

newSubPopSizeExpr: An expression that will be evaluated as an array of new subpopulation sizes

newSubPopSizeFunc: A function that takes parameters `gen` (generation number) and `oldsize` (an array of current population size) and return an array of subpopulation sizes of the next generation. This is usually easier to use than its expression version of this parameter.

subPop: If this parameter is given, the mating scheme will be applied only to the given (virtual) subpopulation. This is only used in heteroMating where mating schemes are passed to.

weight: When subPop is virtual, this is used to detemine the number of offspring for this mating scheme. Weight can be

- 0 (default) the weight will be proportional to the current (virtual) subpopulation size. If other virutal subpopulation has non-zero weight, this virtual subpopulation will produce no offspring (weight 0).

- any negative number -n: the size will be n*m where m is the size of the (virtual) subpopulation of the parental generation.
- any positive number n: the size will be determined by weights from all (virtual) subpopulations.

create a mating scheme (do not use this base mating scheme, use one of its derived classes instead) By default, a mating scheme keeps a constant population size, generates one offspring per mating event. These can be changed using certain parameters. `newSubPopSize`, `newSubPopSizeExpr` and `newSubPopSizeFunc` can be used to specify subpopulation sizes of the offspring generation.

clone()
Deep copy of a mating scheme

submitScratch(*pop, scratch*)
A common submit procedure is defined.

1.2.2 Class `noMating` (Applicable to `all ploidy`)

A mating scheme that does nothing In this scheme, there is

- no mating. Parent generation will be considered as offspring generation.
- no subpopulation change. *During-mating* operators will be applied, but the return values are not checked. I.e., subpopulation size parameters will be ignored although some during-mating operators might be applied.

Note that because the offspring population is the same as parental population, this mating scheme can not be used with other mating schemes in a heterogeneous mating scheme. cloneMating is recommended for that purpose.

class noMating(*numOffspring=1.0, numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeExpr="", newSubPopSizeFunc=None, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)
creat a scheme with no mating

Note All parameters are ignored!
clone()
Deep copy of a scheme with no mating

1.2.3 Class `cloneMating` (Applicable to `all ploidy`)

A clone mating that copy everyone from parental to offspring generation. Note that

- selection is not considered (fitness is ignored)
- sequentialParentMating is used. If offspring (virtual) subpopulation size is smaller than parental subpopulation size, not all parents will be cloned. If offspring (virtual) subpopulation size is larger, some parents will be cloned more than once.
- numOffspring interface is respected.
- during mating operators are applied.

class cloneMating(*numOffspring=1., numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeExpr="", newSubPopSizeFunc=None, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)
create a binomial selection mating scheme Please refer to class `mating` for parameter descriptions.

clone()
Deep copy of a binomial selection mating scheme

1.2.4 Class `binomialSelection` (Applicable to `all ploidy`)

A mating scheme that uses binomial selection, regardless of sex No sex information is involved (binomial random selection). Offspring is chosen from parental generation by random or according to the fitness values. In this mating scheme,

- `numOffspring` protocol is honored;
- population size changes are allowed;
- selection is possible;
- haploid population is allowed.

class binomialSelection(*numOffspring=1., numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeExpr="", newSubPopSizeFunc=None, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)
create a binomial selection mating scheme Please refer to class `mating` for parameter descriptions.

clone()
Deep copy of a binomial selection mating scheme

1.2.5 Class `baseRandomMating` (Applicable to `diploid only`)

This base class defines a general random mating scheme that makes full use of a general random parents chooser, and a Mendelian offspring generator. A general random parents chooser allows selection without replacement, polygemous parents selection (a parent with more than one partners), and the definition of several alpha individuals.Direct use of this mating scheme is not recommended. `randomMating`, `monogemousMating`, `polygemousMating`, `alphaMating` are all special cases of this mating scheme. They should be used whenever possible.

class baseRandomMating(*replacement=True, replenish=False, polySex=Male, polyNum=1, alphaSex=Male, alphaNum=0, alphaField=string, numOffspring=1., numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeExpr="", newSubPopSizeFunc=None, contWhenUniSex=True, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)
alphaField: If an information field is given, individuals with non-zero values at this information field are alpha individuals. Note that these individuals must have `alphaSex`.

alphaNum: Number of alpha individuals. If `infoField` is not given, `alphaNum` random individuals with `alphaSex` will be chosen. If selection is enabled, individuals with higher+ fitness values have higher probability to be selected. There is by default no alpha individual (`alphaNum` = 0).

alphaSex: The sex of the alpha individual, i.e. alpha male or alpha female who be the only mating individuals in their sex group.

polyNum: Number of sex partners.

polySex: Sex of polygamous mating. Male for polygyny, Female for polyandry.

replacement: If set to `True`, a parent can be chosen to mate again. Default to False.

replenish: In case that `replacement=True`, whether or not replenish a sex group when it is exhausted.

FIXME: No document

clone()
Deep copy of a random mating scheme

1.2.6 Class `randomMating` (Applicable to `diploid only`)

A mating scheme of basic sexually random mating In this scheme, sex information is considered for each individual, and ploidy is always 2. Within each subpopulation, males and females are randomly chosen. Then randomly get one copy of chromosomes from father and mother. If only one sex exists in a subpopulation, a parameter (`contWhenUniSex`) can be set to determine the behavior. Default to continuing without warning.

class randomMating(*numOffspring=1., numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeFunc=None, newSubPopSizeExpr="", contWhenUniSex=True, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)
contWhenUniSex: Continue when there is only one sex in the population. Default to `True`.

Please refer to class `mating` for descriptions of other parameters.

clone()
Deep copy of a random mating scheme

1.2.7 Class `selfMating` (Applicable to `diploid only`)

A mating scheme of selfing In this mating scheme, a parent is choosen randomly, acts both as father and mother in the usual random mating. The parent is chosen randomly, regardless of sex. If selection is turned on, the probability that an individual is chosen is proportional to his/her fitness.

class selfMating(*numOffspring=1., numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeFunc=None, newSubPopSizeExpr="", contWhenUniSex=True, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)
contWhenUniSex: Continue when there is only one sex in the population. Default to `True`.

create a self mating scheme Please refer to class `mating` for descriptions of other parameters.

clone()
Deep copy of a self mating scheme

1.2.8 Class `monogamousMating` (Applicable to `diploid only`)

A mating scheme of monogamy This mating scheme is identical to random mating except that parents are chosen without replacement. Under this mating scheme, offspring share the same mother must share the same father. In case that all parental pairs are exhausted, parameter `replenish=True` allows for the replenishment of one or both sex groups.

class monogamousMating(*replenish=False, numOffspring=1., numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeFunc=None, newSubPopSizeExpr="", contWhenUniSex=True, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)
`replenish` This parameter allows replenishment of one or both parental sex groups in case that they are are exhausted. Default to False. Please refer to class `mating` for descriptions of other parameters.

clone()
Deep copy of a random mating scheme

1.2.9 Class `polygamousMating` (Applicable to `diploid only`)

A mating scheme of polygymy or polyandry This mating scheme is composed of a random parents chooser that allows for polygamous mating, and a mendelian offspring generator. In this mating scheme, a male (or female) parent will have more than one sex partner (`numPartner`). Parents returned from this parents chooser will yield the same male (or female) parents, each with varying partners.

class `polygamousMating`(*polySex=Male, polyNum=1, replacement=False, replenish=False, numOffspring=1., numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeFunc=None, newSubPopSizeExpr="", contWhenUniSex=True, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)

polyNum: Number of sex partners.

polySex: Sex of polygamous mating. Male for polygyny, Female for polyandry.

replacement: If set to `True`, a parent can be chosen to mate again. Default to False.

replenish: In case that `replacement=True`, whether or not replenish a sex group when it is exhausted. Please refer to class `mating` for descriptions of other parameters.

FIXME: No document

`clone`()

Deep copy of a random mating scheme

1.2.10 Class `consanguineousMating` (Applicable to `diploid only`)

A mating scheme of consanguineous mating In this mating scheme, a parent is choosen randomly and mate with a relative that has been located and written to a number of information fields.

class `consanguineousMating`(*relativeFields=[], func=None, param=None, replacement=False, replenish=True, numOffspring=1., numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeFunc=None, newSubPopSizeExpr="", contWhenUniSex=True, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)

func: A python function that can be used to prepare the indexes of these information fields. For example, functions population::locateRelatives and/or population::setIndexesOfRelatives can be used to locate certain types of relatives of each individual.

param: An optional parameter that can be passed to `func`.

relativeFields: The information fields that stores indexes to other individuals in a population. If more than one valid (positive value) indexes exist, a random index will be chosen. (c.f. `infoParentsChooser`) If there is no individual having any valid index, the second parent will be chosen randomly from the whole population.

create a consanguineous mating scheme This mating scheme randomly choose a parent and then choose his/her spouse from indexes stored in `infoFields`. Please refer to `infoParentsChooser` and `mendelianOffspringGenerator` for other parameters.

`clone`()

Deep copy of a consanguineous mating scheme

1.2.11 Class `alphaMating` (Applicable to `diploid only`)

Only a number of alpha individuals can mate with individuals of opposite sex. This mating scheme is composed of an random parents chooser with alpha individuals, and a Mendelian offspring generator. That is to say, a certain number

of alpha individual (male or female) are determined by `alphaNum` or an information field. Then, only these alpha individuals are able to mate with random individuals of opposite sex.

class alphaMating(*alphaSex=Male, alphaNum=0, alphaField=string, numOffspring=1., numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeFunc=None, newSubPopSizeExpr="", subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)

alphaField: If an information field is given, individuals with non-zero values at this information field are alpha individuals. Note that these individuals must have `alphaSex`.

alphaNum: Number of alpha individuals. If `infoField` is not given, `alphaNum` random individuals with `alphaSex` will be chosen. If selection is enabled, individuals with higher+ fitness values have higher probability to be selected. There is by default no alpha individual (`alphaNum` = 0).

alphaSex: The sex of the alpha individual, i.e. alpha male or alpha female who be the only mating individuals in their sex group.

Please refer to class `mating` for descriptions of other parameters. Note: If selection is enabled, it works regularly on on-alpha sex, but works twice on alpha sex. That is to say, `alphaNum` alpha indiviudals are chosen selectively, and selected again during mating.

clone()

Deep copy of a random mating scheme

1.2.12 Class `haplodiploidMating` (Applicable to `haplodiploid only`)

Haplodiploid mating scheme of many hymemopterans This mating scheme is composed of an alphaParentChooser and a haplodiploidOffspringGenerator. The alphaParentChooser chooses a single Female randomly or from a given information field. This female will mate with random males from the colony. The offspring will have one of the two copies of chromosomes from the female parent, and the first copy of chromosomes from the male parent. Note that if a recombinator is used, it should disable recombination of male parent.

class haplodiploidMating(*alphaSex=Female, alphaNum=1, alphaField=string, numOffspring=1., numOffspringFunc=None, maxNumOffspring=0, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex, newSubPopSize=[], newSubPopSizeFunc=None, newSubPopSizeExpr="", subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)

alphaField: Information field that identifies the queen of the colony. By default, a random female will be chosen.

alphaNum: Number of alpha individual. Default to one.

alphaSex: Sex of the alpha individual. Default to Female.

Please refer to class `mating` for descriptions of other parameters.

clone()

Deep copy of a random mating scheme

1.2.13 Class `pedigreeMating` (Applicable to `all ploid`)

A mating scheme that follows a given pedigree In this scheme, a pedigree is given and the mating scheme will choose parents and produce offspring strictly following the pedigree. Parameters setting number of offspring per mating event, and size of the offspring generations are ignored.To implement this mating scheme in pyMating, 1.) a newSubPopSizeFunc should be given to return the exact subpopulation size, returned from pedigree.subPopSizes(gen). 2.) use pedigreeChooser to choose parents 3.) use a suitable offspring generator to generate offspring.This pedigreeMating helps you do 1 and 2, and use a mendelianOffspringGenerator as the default offspring generator. You can use another

offspring generator by setting the generator parameter. Note that the offspring generator can generate one and only one offspring each time.

class pedigreeMating(*ped, generator, newSubPopSize=[], newSubPopSizeFunc=None, newSubPopSizeExpr="", subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)

Please refer to class `mating` for descriptions of other parameters.

clone()

Deep copy of a random mating scheme

1.2.14 Class `pyMating` (Applicable to `all ploidy`)

A Python mating scheme This hybrid mating scheme does not have to involve a python function. It requires a parent chooser, and an offspring generator. The parent chooser chooses parent(s) and pass them to the offspring generator to produce offspring.

class pyMating(*chooser, generator, newSubPopSize=[], newSubPopSizeExpr="", newSubPopSizeFunc=None, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)

chooser: A parent chooser that chooses parent(s) from the parental generation.

generator: An offspring generator that produce offspring of given parents.

create a Python mating scheme

clone()

Deep copy of a Python mating scheme

1.2.15 Class `heteroMating` (Applicable to `diploid only`)

A heterogeneous mating scheme that applies a list of mating schemes to different (virtual) subpopulations.

class heteroMating(*matingSchemes, newSubPopSize=[], newSubPopSizeExpr="", newSubPopSizeFunc=None, shuffleOffspring=True, subPop=InvalidSubPopID, virtualSubPop=InvalidSubPopID, weight=0*)

matingSchemes: A list of mating schemes. If parameter `subPop` of an mating scheme is specified, it will be applied to specific subpopulation. If `virtualSubPop` if specified, it will be applied to specifc virtual subpopulations.

create a heterogeneous Python mating scheme Parameter subpop, virtualSubPOp and weight of this mating scheme is ignored.

clone()

Deep copy of a Python mating scheme

1.2.16 Class `sequentialParentChooser` (Applicable to `all ploidy`)

This parent chooser chooses a parent linearly, regardless of sex or fitness values (selection is not considered).

class sequentialParentChooser()

FIXME: No document

clone()

FIXME: No document

1.2.17 Class `sequentialParentsChooser` (Applicable to `all ploidy`)

This parents chooser chooses two parents sequentially. The parents are chosen from their respective sex groups. Selection is not considered.

class `sequentialParentsChooser`()
FIXME: No document

`clone`()
FIXME: No document

1.2.18 Class `randomParentChooser` (Applicable to `all ploidy`)

This parent chooser chooses a parent randomly from the parental generation. If selection is turned on, parents are chosen with probabilities that are proportional to their fitness values. Sex is not considered. Parameter `replacement` determines if a parent can be chosen multiple times. In case that `replacement=false`, paremeter `replenish=true` allows restart of the process if all parents are exhausted. Note that selection is not allowed when `replacement=false` because this poses a particular order on individuals in the offspring generation.

class `randomParentChooser`(*replacement=True, replenish=False*)
replacement: If replacement is false, a parent can not be chosen more than once.

replenish: If all parent has been chosen, choose from the whole parental population again.

FIXME: No document

`clone`()
FIXME: No document

1.2.19 Class `randomParentsChooser` (Applicable to `all ploidy`)

This parent chooser chooses two parents randomly, a male and a female, from their respective sex groups randomly. If selection is turned on, parents are chosen from their sex groups with probabilities that are proportional to their fitness values. If parameter `replacement` is false, a chosen pair of parents can no longer be selected. This feature can be used to simulate monopoly. If `replenish` is true, a sex group can be replenished when it is exhausted. Note that selection is not allowed in the case of monopoly because this poses a particular order on individuals in the offspring generation. This parents chooser also allows polygamous mating by reusing a parent multiple times when returning parents, and allows specification of a few alpha individuals who will be the only mating individuals in their sex group.

class `randomParentsChooser`(*replacement=True, replenish=False, polySex=Male, polyNum=1, alphaSex=Male, alphaNum=0, alphaField=string*)
alphaField: If an information field is given, individuals with non-zero values at this information field are alpha individuals. Note that these individuals must have `alphaSex`.

alphaNum: Number of alpha individuals. If `infoField` is not given, `alphaNum` random individuals with `alphaSex` will be chosen. If selection is enabled, individuals with higher fitness values have higher probability to be selected. There is by default no alpha individual (`alphaNum` = 0).

alphaSex: The sex of the alpha individual, i.e. alpha male or alpha female who be the only mating individuals in their sex group.

polyNum: Number of sex partners.

polySex: Male (polygyny) or Female (polyandry) parent that will have `polyNum` sex partners.

replacement: Choose with (`True`, default) or without (`False`) replacement. When choosing without replacement, parents will be paired and can only mate once.

replenish: If set to true, one or both sex groups will be replenished if they are exhausted.

Note: If selection is enabled, it works regularly on on-alpha sex, but works twice on alpha sex. That is to say, `alphaNum` alpha indiviudals are chosen selectively, and selected again during mating.

clone()
FIXME: No document

1.2.20 Class `infoParentsChooser` (Applicable to `all ploidy`)

This parents chooser choose an individual randomly, but choose his/her spouse from a given set of information fields, which stores indexes of individuals in the same generation. A field will be ignored if its value is negative, or if sex is compatible.Depending on what indexes are stored in these information fields, this parent chooser can be used to implement consanguineous mating where close relatives are located for each individual, or certain non-random mating schemes where each individual can only mate with a small number of pre-determinable individuals.This parent chooser (currently) uses `randomParentChooser` to choose one parent and randomly choose another one from the information fields. Because of potentially non-even distribution of valid information fields, the overall process may not be as random as expected, especially when selection is applied.Note: if there is no valid individual, this parents chooser works like a double parentChooser.

class infoParentsChooser(*infoFields=[], replacement=True, replenish=False*)
infoFields: Information fields that store index of matable individuals.

replacement: If replacement is false, a parent can not be chosen more than once.

replenish: If all parent has been chosen, choose from the whole parental population again.

FIXME: No document

clone()
FIXME: No document

1.2.21 Class `pedigreeParentsChooser` (Applicable to `all ploidy`)

This parents chooser chooses one or two parents from a given pedigree. It works even when only one parent is needed.

class pedigreeParentsChooser(*ped*)
FIXME: No document

clone()
FIXME: No document

subPopSizes(*gen*)
FIXME: No document

1.2.22 Class `pyParentsChooser` (Applicable to `all ploidy`)

This parents chooser accept a Python generator function that yields repeatedly an index (relative to each subpopulation) of a parent, or indexes of two parents as a Python list of tuple. The generator function is responsible for handling sex or selection if needed.

class pyParentsChooser(*parentsGenerator*)
parentsGenerator: A Python generator function

FIXME: No document

clone()
FIXME: No document

finalize(*pop, sp*)
FIXME: No document

1.2.23 Class `cloneOffspringGenerator` (Applicable to `all ploidy`)

Clone offspring generator copies parental geneotype to a number of offspring. Only one parent is accepted. The number of offspring produced is controled by parameters `numOffspring`, `numOffspringFunc`, `maxNumOffspring` and `mode`. Parameters `sexParam` and `sexMode` is ignored.

class cloneOffspringGenerator(*numOffspring=1, numOffspringFunc=None, maxNumOffspring=1, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex*)
sexMode: Ignored because sex is copied from the parent.

sexParam: Ignored because sex is copied from the parent.

FIXME: No document

clone()
FIXME: No document

1.2.24 Class `selfingOffspringGenerator` (Applicable to `diploid only`)

Selfing offspring generator works similarly as a mendelian offspring generator but a single parent produces both the paternal and maternal copy of the offspring chromosomes. This offspring generator accepts a dipload parent. A random copy of the parental chromosomes is chosen randomly to form the parental copy of the offspring chromosome, and is chosen randomly again to form the maternal copy of the offspring chromosome.

class selfingOffspringGenerator(*numOffspring=1, numOffspringFunc=None, maxNumOffspring=1, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex*)
FIXME: No document

clone()
FIXME: No document

1.2.25 Class `haplodiploidOffspringGenerator` (Applicable to `haplodiploid only`)

Haplodiploid offspring generator mimics sex-determination in honey bees. Given a female (queen) parent and a male parent, the female is considered as diploid with two set of chromosomes, and the male is condiered as haploid. Actually, the first set of male chromosomes are used. During mating, female produce eggs, subject to potential recombination and gene conversion, while male sperm is identical to the parental chromosome.Female offspring has two sets of chromosomes, one from mother and one from father. Male offspring has one set of chromosomes from his mother.

class haplodiploidOffspringGenerator(*numOffspring=1, numOffspringFunc=None, maxNumOffspring=1, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex*)
FIXME: No document

clone()
FIXME: No document

copyParentalGenotype(*parent, it, ploidy, count*)
FIXME: No document

1.2.26 Class `mendelianOffspringGenerator` (Applicable to `diploid only`)

Mendelian offspring generator accepts two parents and pass their genotype to a number of offspring following Mendelian's law. Basically, one of the paternal chromosomes is chosen randomly to form the paternal copy of the offspring, and one of the maternal chromosome is chosen randomly to form the maternal copy of the offspring. The number of offspring produced is controled by parameters `numOffspring`, `numOffspringFunc`, `maxNumOffspring` and `mode`. Recombination will not happen unless a during-mating operator recombinator is used.

class `mendelianOffspringGenerator`(*numOffspring=1, numOffspringFunc=None, maxNumOffspring=1, mode=MATE_NumOffspring, sexParam=0.5, sexMode=MATE_RandomSex*)

FIXME: No document

`clone`()

FIXME: No document

`formOffspringGenotype`(*parent, it, ploidy, count*)

Does not set sex if count == -1.

count: Index of offspring, used to set offspring sex

1.3 Simulator

1.3.1 Class `simulator`

Simulator manages several replicates of a population, evolve them using given mating scheme and operators Simulators combine three important components of simuPOP: population, mating scheme and operator together. A simulator is created with an instance of `population`, a replicate number `rep` and a mating scheme. It makes `rep` number of replicates of this population and control the evolutionary process of them.

The most important function of a simulator is `evolve()`. It accepts an array of operators as its parameters, among which, `preOps` and `postOps` will be applied to the populations at the beginning and the end of evolution, respectively, whereas `ops` will be applied at every generation.

A simulator separates operators into *pre-*, *during-*, and *post-mating* operators. During evolution, a simulator first apply all pre-mating operators and then call the `mate()` function of the given mating scheme, which will call during-mating operators during the birth of each offspring. After mating is completed, post-mating operators are applied to the offspring in the order at which they appear in the operator list.

Simulators can evolve a given number of generations (the `end` parameter of `evolve`), or evolve indefinitely until a certain type of operators called terminator terminates it. In this case, one or more terminators will check the status of evolution and determine if the simulation should be stopped. An obvious example of such a terminator is a fixation-checker.

A simulator can be saved to a file in the format of `'txt'`, `'bin'`, or `'xml'`. This allows you to stop a simulator and resume it at another time or on another machine.

class `simulator`(*pop, matingScheme, stopIfOneRepStops=False, applyOpToStoppedReps=False, rep=1*)

applyOpToStoppedReps: If set, the simulator will continue to apply operators to all stopped replicates until all replicates are marked 'stopped'.

matingScheme: A mating scheme

population: A population created by `population()` function. This population will be copied `rep` times to the simulator. Its content will not be changed.

rep: Number of replicates. Default to `1`.

stopIfOneRepStops: If set, the simulator will stop evolution if one replicate stops.

create a simulator

addInfoField(*field, init=0*)

Add an information field to all replicates Add an information field to all replicate, and to the simulator itself. This is important because all populations must have the same genotypic information as the simulator. Adding an information field to one or more of the replicates will compromise the integrity of the simulator.

field: Information field to be added

addInfoFields(*fields, init=0*)

Add information fields to all replicates Add given information fields to all replicate, and to the simulator itself.

clone()

Deep copy of a simulator

evolve(*ops, preOps=[], postOps=[], end=-1, gen=-1, dryrun=False*)

Evolve all replicates of the population, subject to operators Evolve to the `end` generation unless `end=-1`. An operator (terminator) may stop the evolution earlier.

`ops` will be applied to each replicate of the population in the order of:

- all pre-mating opertors
- during-mating operators called by the mating scheme at the birth of each offspring
- all post-mating operators If any pre- or post-mating operator fails to apply, that replicate will be stopped. The behavior of the simulator will be determined by flags `applyOpToStoppedReps` and `stopIfOneRepStopss`.

dryrun: Dryrun mode. Default to `False`.

gen: Generations to evolve. Default to `-1`. In this case, there is no ending generation and a simulator will only be ended by a terminator. Note that simu.gen() refers to the begining of a generation, and starts at 0.

ops: Operators that will be applied at each generation, if they are active at that generation. (Determined by the `begin`, `end`, `step` and `at` parameters of the operator.)

postOps: Operators that will be applied after evolution. `evolve()` function will *not* check if they are active.

preOps: Operators that will be applied before evolution. `evolve()` function will *not* check if they are active.

Note: When `gen = -1`, you can not specify negative generation parameters to operators. How would an operator know which genertion is the -1 genertion if no ending genertion is given?

gen()

Return the current generation number

getPopulation(*rep, destructive=False*)

Return a copy of population `rep`By default return a cloned copy of population `rep` of the simulator. If `destructive==True`, the population is extracted from the simulator, leaving a defunct simulator.

destructive: If true, destroy the copy of population within this simulator. Default to false. `getPopulation(rep, true)` is a more efficient way to get hold of a population when the simulator will no longer be used.

rep: The index number of the replicate which will be obtained

numRep()

Return the number of replicates

population(*rep*)

Return a reference to the `rep` replicate of this simulator.

rep: The index number of replicate which will be accessed

Note: The returned reference is temporary in the sense that the refered population will be invalid after another round of evolution. If you would like to get a persistent population, please use `getPopulation(rep)`.

saveSimulator (*filename, format="", compress=True*)
Save simulator in `'txt'`, `'bin'` or `'xml'` format

compress: Obsolete parameter

filename: Filename to save the simulator. Default to `simu`.

format: Obsolete parameter

setAncestralDepth (*depth*)
Set ancestral depth of all replicates

setGen (*gen*)
Set the current generation. Usually used to reset a simulator.

gen: New generation index number

setMatingScheme (*matingScheme*)
Set a new mating scheme

step (*ops=[], preOps=[], postOps=[], steps=1, dryrun=False*)
Evolve `steps` generation

vars (*rep, subPop=-1*)
Return the local namespace of population `rep`, equivalent to `x.population(rep).vars(subPop)`.

1.4 Pedigree

1.4.1 Class `pedigree`

A pedigree manipulation class. A pedigree has all the pedigree information that is needed to look at parent offspring relationship in a multi-generation population.Conceptually, there are n generations with the latest generation being generation 0. The number of generations (c.f. gen()) is the number of parental generations plus 1. Therefore, each individual can be identified by (gen, idx).Each individual can have a few properties 1. mother (c.f. mother()) 2. father (c.f. father(), optional because a pedigree can have only one sex) 3. subpopulation (if subpopulation structure is given) 4. sex (c.f. info('sex')) 5. affection (c.f. info('affection') 6. arbitrary information fields (c.f. info())

class pedigree (*numParents=2, pedfile=string*)
will be loaded from this file.

addGen (*subPopSize*)
For the new generation.

addInfo (*name, init=0*)
Add an information field to the pedigree, with given initial value

clone ()
Make a copy of this pedigree.

father (*gen, idx*)
The returned index is the absolute index of father in the parental generation.

gen ()
Return the number of generations of this pedigree.

info (*gen, idx, name*)
Return information `name` of individual `idx` at generation `gen`.

info (*gen, name*)
Return information `name` of all individuals at generation `gen`.

info (*gen, subPop, idx, name*)
Return information `name` of individual `idx` of subpopulation `subPop` at generation `gen`.

load(*filename*)
PARENTSTAGGER. The format is described in the simuPOP reference manual

loadInfo(*filename, name*)
AFFECTIONTAGGER, pyTagger and infoTagger.

markUnrelated()
Last generation (not marked by selectIndividuals) from the pedigree.

mother(*gen, idx*)
The returned index is the absolute index of mother in the parental generation.

numParents()
Return the number of parents for each individual

popSize(*gen*)
Population size at generation gen

removeUnrelated(*adjust_index=True*)
WARNING: if adjust_index=false, an invalid pedigree will be generated.

save(*filename*)
Write the pedigree to a file.

saveInfo(*filename, name*)
Save auxiliary information name to an information pedigree file.

saveInfo(*filename, names*)
Save auxiliary information names to an information pedigree file

selectIndividuals(*inds*)
REMOVEUNRELATED function will remove these individuals from the pdeigree

setFather(*parent, gen, idx*)
Set the index of the father of individual idx at generation gen.

setFather(*parent, gen, subPop, idx*)
Set the index of the father of individual idx of subpopulation subPop at generation gen.

setInfo(*info, gen, idx, name*)
Set information name of individual idx at generation gen.

setInfo(*info, gen, subPop, idx, name*)
Set information name of individual idx of subpopulation subPop at generation gen.

setMother(*parent, gen, idx*)
Set the index of the mother of individual idx at generation gen.

setMother(*parent, gen, subPop, idx*)
Set the index of the mother of individual idx of subpopulation subPop at generation gen.

subPopSize(*gen, subPop*)
Return the subpopulation size of subpopulation subPop of generation gen.

subPopSizes(*gen*)
Return the subpopulation sizes of generation gen.

CHAPTER

TWO

Operator References

This chapter will list all functions, types and operators by category. The reference for `class baseOperator` is in section **??**.

2.1 The common interface of operators

2.1.1 Class `baseOperator`

Base class of all classes that manipulate populations Operators are objects that act on populations. They can be applied to populations directly using their function forms, but they are usually managed and applied by a simulator.

There are three kinds of operators:

- built-in: written in C++, the fastest. They do not interact with Python shell except that some of them set variables that are accessible from Python.

- hybrid: written in C++ but calls a Python function during execution. Less efficient. For example, a hybrid mutator `pyMutator` will go through a population and mutate alleles with given mutation rate. How exactly the allele will be mutated is determined by a user-provided Python function. More specifically, this operator will pass the current allele to a user-provided Python function and take its return value as the mutant allele.

- pure Python: written in Python. The same speed as Python. For example, a `varPlotter` can plot Python variables that are set by other operators. Usually, an individual or a population object is passed to a user-provided Python function. Because arbitrary operations can be performed on the passed object, this operator is very flexible.

Operators can be applied at different stages of the life cycle of a generation. It is possible for an operator to apply multiple times in a life cycle. For example, a `savePopulation` operator might be applied before and after mating to trace parental information. More specifically, operators can be applied at *pre-*, *during-*, *post-mating*, or a combination of these stages. Applicable stages are usually set by default but you can change it by setting `stage=(PreMating|PostMating|DuringMating|PrePostMating|PreDuringMating|DuringPostMating)` parameter. Some operators ignore `stage` parameter because they only work at one stage.

Operators do not have to be applied at all generations. You can specify starting and/or ending generations (parameters `start`, `end`), gaps between applicable generations (parameter `step`), or specific generations (parameter `at`). For example, you might want to start applying migrations after certain burn-in generations, or calculate certain statistics only sparsely. Generation numbers can be counted from the last generation, using negative generation numbers.

Most operators are applied to every replicate of a simulator during evolution.Operators can have outputs, which can be standard (terminal) or a file. Output can vary with replicates and/or generations, and outputs from different operators can be accumulated to the same file to form table-like outputs.

Filenames can have the following format:

- `'filename'` this file will be overwritten each time. If two operators output to the same file, only the last one will succeed;
- `'>filename'` the same as `'filename'`;
- `'>>filename'` the file will be created at the beginning of evolution (`simulator::evolve`) and closed at the end. Outputs from several operators are appended;
- `'>>>filename'` the same as `'>>filename'` except that the file will not be cleared at the beginning of evolution if it is not empty;
- `'>'` standard output (terminal);
- `''` suppress output.

The output filename does not have to be fixed. If parameter `outputExpr` is used (parameter `output` will be ignored), it will be evaluated when a filename is needed. This is useful when you need to write different files for different replicates/generations.

class `baseOperator`(*output, outputExpr, stage, begin, end, step, at, rep, infoFields*)

at: An array of active generations. If given, `stage`, `begin`, `end`, and `step` will be ignored.

begin: The starting generation. Default to `0`. A negative number is allowed.

end: Stop applying after this generation. A negative numbers is allowed.

output: A string of the output filename. Different operators will have different default `output` (most commonly `'>'` or `''`).

outputExpr: An expression that determines the output filename dynamically. This expression will be evaluated against a population's local namespace each time when an output filename is required. For example, `"'>>out%s_%s.xml' % (gen, rep)"` will output to `>>>out1_1.xml` for replicate `1` at generation `1`.

rep: Applicable replicates. It can be a valid replicate number, `REP_ALL` (all replicates, default), or `REP_LAST` (only the last replicate). `REP_LAST` is useful in adding newlines to a table output.

step: The number of generations between active generations. Default to `1`.

common interface for all operators (this base operator does nothing by itself.)

Note

- Negative generation numbers are allowed for parameters `begin`, `end` and `at`. They are interpreted as `endGen + gen + 1`. For example, `begin = -2` in `simu.evolve(..., end=20)` starts at generation `19`.
- `REP_ALL, REP_LAST` are special constant that can only be used in the constructor of an operator. That is to say, explicit test of `rep() == REP_LAST` will not work.

`apply`(*pop*)

Apply to one population. It does not check if the operator is activated.

`clone`()

Deep copy of an operator

`diploidOnly`()

Determine if the operator can be applied only for diploid population

haploidOnly()
Determine if the operator can be applied only for haploid population

infoField(*idx*)
Get the information field specified by user (or by default)

infoSize()
Get the length of information fields for this operator

2.2 Initialization

2.2.1 Class `initializer`

Initialize alleles at the start of a generation Initializers are used to initialize populations before evolution. They are set to be `PreMating` operators by default. simuPOP provides three initializers. One assigns alleles by random, one assigns a fixed set of genotypes, and the last one calls a user-defined function.

class initializer(*subPop=[], indRange=[], loci=[], atPloidy=-1, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
atPloidy: Initialize which copy of chromosomes. Default to all.

indRange: A `[begin, end]` pair of the range of absolute indexes of individuals, for example, `([1,2])`; or an array of `[begin, end]` pairs, such as `([[1,4],[5,6]])`. This is how you can initialize individuals differently within subpopulations. Note that ranges are in the form of [a,b). I.e., range [4,6] will intialize individual 4, 5, but not 6. As a shortcut for [4,5], you can use [4] to specify one individual.

loci: A vector of locus indexes at which initialization will be done. If empty, apply to all loci.

locus: A shortcut to `loci`

subPop: An array specifies applicable subpopulations

create an initializer. Default to be always active.

clone()
Deep copy of an initializer

2.2.2 Class `initSex` (Function form: `InitSex`)

An operator to initialize individual sex. For convenience, this operator is included by other initializers such as initByFreq, initByValue, or pyInit.

class initSex(*maleFreq=0.5, sex=[], subPop=[], indRange=[], loci=[], atPloidy=-1, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
maleFreq: Male frequency. Default to `0.5`. Sex will be initialized with this parameter.

sex: A list of sexes (Male or Female) and will be applied to individuals in in turn. If specified, parameter `maleFreq` is ignored.

initialize individual sex.

apply(*pop*)
Apply this operator to population `pop`

clone()
Deep copy of an initSex

2.2.3 Class `initByFreq` (Function form: `InitByFreq`)

Initialize genotypes by given allele frequencies, and sex by male frequency This operator assigns alleles at `loci` with given allele frequencies. By default, all individuals will be assigned with random alleles. If `identicalInds=True`, an individual is assigned with random alleles and is then copied to all others. If `subPop` or `indRange` is given, multiple arrays of `alleleFreq` can be given to given different frequencies for different subpopulation or individual ranges.

class `initByFreq`(*alleleFreq=[], identicalInds=False, subPop=[], indRange=[], loci=[], atPloidy=-1, maleFreq=0.5, sex=[], stage=PreMating, begin=0, end=1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

alleleFreq: An array of allele frequencies. The sum of all frequencies must be 1; or for a matrix of allele frequencies, each row corresponses to a subpopulation or range.

identicalInds: Whether or not make individual genotypes identical in all subpopulations. If `True`, this operator will randomly generate genotype for an individual and spread it to the whole subpopulation in the given range.

sex: An array of sex `[Male, Female, Male...]` for individuals. The length of sex will not be checked. If it is shorter than the number of individuals, sex will be reused from the beginning.

stage: Default to `PreMating`.

randomly assign alleles according to given allele frequencies

`apply`(*pop*)
Apply this operator to population `pop`

`clone`()
Deep copy of the operator `initByFreq`

2.2.4 Class `initByValue` (Function form: `InitByValue`)

Initialize genotype by value and then copy to all individuals Operator `initByValue` gets one copy of chromosomes or the whole genotype (or of those corresponds to `loci`) of an individual and copy them to all or a subset of individuals. This operator assigns given alleles to specified individuals. Every individual will have the same genotype. The parameter combinations should be

- `value - subPop/indRange`: individual in `subPop` or in range(s) will be assigned genotype `value`;
- `subPop/indRange`: `subPop` or `indRange` should have the same length as `value`. Each item of `value` will be assigned to each `subPop` or `indRange`.

class `initByValue`(*value=[], loci=[], atPloidy=-1, subPop=[], indRange=[], proportions=[], maleFreq=0.5, sex=[], stage=PreMating, begin=0, end=1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

maleFreq: Male frequency

proportions: An array of percentages for each item in `value`. If given, assign given genotypes randomly.

sex: An array of sex `[Male, Female, Male...]` for individuals. The length of sex will not be checked. If length of sex is shorter than the number of individuals, sex will be reused from the beginning.

stages: Default to `PreMating`.

value: An array of genotypes of one individual, having the same length as the length of `loci()` or `loci()*ploidy()` or `pop.genoSize()` (whole genotype) or `totNumLoci()` (one copy of chromosomes). This parameter can also be an array of arrays of genotypes of one individual. If `value` is an array of values, it should have the length one, number of subpopulations, or the length of ranges of proportions.

initialize a population by given alleles

apply(*pop*)
Apply this operator to population `pop`

clone()
Deep copy of the operator `initByValue`

2.2.5 Class `spread` (Function form: `Spread`)

Copy the genotype of an individual to all individuals Function `Spread(ind, subPop)` spreads the genotypes of `ind` to all individuals in an array of subpopulations. The default value of `subPop` is the subpopulation where `ind` resides.

class spread(*ind, subPop=[], stage=PreMating, begin=0, end=1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
copy genotypes of `ind` to all individuals in `subPop`

apply(*pop*)
Apply this operator to population `pop`

clone()
Deep copy of the operator `spread`

2.2.6 Class `pyInit` (Function form: `PyInit`)

A python operator that uses a user-defined function to initialize individuals. This is a hybrid initializer. Users of this operator must supply a Python function with parameters allele, ploidy and subpopulation indexes `(index, ploidy, subPop)`, and return an allele value. This operator will loop through all individuals in each subpopulation and call this function to initialize populations. The arrange of parameters allows different initialization scheme for each subpopulation.

class pyInit(*func, subPop=[], loci=[], atPloidy=-1, indRange=[], maleFreq=0.5, sex=[], stage=PreMating, begin=0, end=1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
atPloidy: Initialize which copy of chromosomes. Default to all.

func: A Python function with parameter `(index, ploidy, subPop)`, where

- `index` is the allele index ranging from `0` to `totNumLoci-1`;
- `ploidy` is the index of the copy of chromosomes;
- `subPop` is the subpopulation index.

The return value of this function should be an integer.

loci: A vector of locus indexes. If empty, apply to all loci.

locus: A shortcut to `loci`.

stage: Default to `PreMating`.

initialize populations using given user function

apply(*pop*)
Apply this operator to population `pop`

clone()
Deep copy of the operator `pyInit`

2.3 Migration

2.3.1 Class `migrator`

Migrate individuals from (virtual) subpopulations to other subpopulations Migrator is the only way to mix genotypes of several subpopulations because mating is strictly within subpopulations in simuPOP. Migrators are quite flexible in simuPOP in the sense that

- migration can happen from and to a subset of subpopulations.
- migration can be done by probability, proportion or by counts. In the case of probability, if the migration rate from subpopulation `a` to `b` is `r`, then everyone in subpopulation `a` will have this probability to migrate to `b`. In the case of proportion, exactly `r*size_of_subPop_a` individuals (chosen by random) will migrate to subpopulation `b`. In the last case, a given number of individuals will migrate.
- new subpopulation can be generated through migration. You simply need to migrate to a subpopulation with a new subpopulation number.

class `migrator` (*rate, mode=MigrByProbability, fromSubPop=[], toSubPop=[], stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

fromSubPop: An array of 'from' subpopulations (a number) or virtual subpopulations (a pair of numbers). Default to all subpopulations. For example, if you define a virtual subpopulation by sex, you can use `fromSubpop=[(0,0), 1]` to choose migrants from the first virtual subpopulation of subpopulation 0, and from subpopulation 1. If a single number `sp` is given, it is intepretted as [sp]. Note that `fromSubPop=(0, 1)` (two subpopulation) is different from `fromSubPop=[(0,1)]` (a virtual subpopulation).

mode: One of `MigrByProbability` (default), `MigrByProportion` or `MigrByCounts`

rate: Migration rate, can be a proportion or counted number. Determined by parameter `mode`. `rate` should be an m by n matrix. If a number is given, the migration rate will be a `m` by `n` matrix of value `r`

stage: Default to `PreMating`

toSubPop: An array of 'to' subpopulations. Default to all subpopulations. If a single subpopulation is specified, `[]` can be ignored.

create a migrator

Note

- The overall population size will not be changed. (Mating schemes can do that). If you would like to keep the subpopulation sizes after migration, you can use the `newSubPopSize` or `newSubPopSizeExpr` parameter of a mating scheme.
- `rate` is a matrix with dimensions determined by `fromSubPop` and `toSubPop`. By default, `rate` is a matrix with element `r(i, j)`, where `r(i, j)` is the migration rate, probability or count from subpopulation `i` to `j`. If `fromSubPop` and/or `toSubPop` are given, migration will only happen between these subpopulations. An extreme case is 'point migration', `rate=[[r]], fromSubPop=a, toSubPop=b` which migrate from subpopulation `a` to `b` with given rate `r`.

`apply` (*pop*)
Apply the `migrator`

`clone`()
Deep copy of a migrator

`rate`()
Return migration rate

setRates(*rate, mode*)
Set migration rate Format should be `0-0 0-1 0-2, 1-0 1-1 1-2, 2-0, 2-1, 2-2`. For mode `MigrByProbability` or `MigrByProportion`, `0-0,1-1,2-2` will be set automatically regardless of input.

2.3.2 Class `pyMigrator`

A more flexible Python migrator This migrator can be used in two ways

- define a function that accepts a generation number and returns a migration rate matrix. This can be used in various migration rate cases.
- define a function that accepts individuals etc, and returns the new subpopulation ID.

More specifically, `func` can be

- `func(ind)` when neither `loci` nor `param` is given.
- `func(ind, genotype)` when `loci` is given.
- `func(ind, param)` when `param` is given.
- `func(ind, genotype, param)` when both `loci` and `param` are given.

class pyMigrator(*rateFunc=None, indFunc=None, mode=MigrByProbability, fromSubPop=[], toSubPop=[], loci=[], param=None, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
indFunc: A Python function that accepts an individual, optional genotypes and parameters, then returns a subpopulation ID. This method can be used to separate a population according to individual genotype.

rateFunc: A Python function that accepts a generation number, current subpopulation sizes, and returns a migration rate matrix. The migrator then migrate like a usual migrator.

stage: Default to `PreMating`

create a hybrid migrator

apply(*pop*)
Apply a `pyMigrator`

clone()
Deep copy of a `pyMigrator`

2.3.3 Class `splitSubPop` (Function form: `SplitSubPop`)

Split a subpopulation

class splitSubPop(*which=0, sizes=[], proportions=[], subPopID=[], randomize=True, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
proportions: Proportions of new subpopulations. Should be added up to `1`.

randomize: Whether or not randomize individuals before population split. Default to True.

sizes: New subpopulation sizes. The sizes should be added up to the original subpopulation (subpopulation `which`) size.

subPopID: New subpopulation IDs. Otherwise, the operator will automatically set new subpopulation IDs to new subpopulations.

which: Which subpopulation to split. If there is no subpopulation structure, use `0` as the first (and only) subpopulation.

split a subpopulation Split a subpopulation by sizes or proportions. Individuals are randomly (by default) assigned to the resulting subpopulations. Because mating schemes may introduce certain order to individuals, randomization ensures that split subpopulations have roughly even distribution of genotypes.

apply(*pop*)
Apply a `splitSubPop` operator

clone()
Deep copy of a `splitSubPop` operator

2.3.4 Class `mergeSubPops` (Function form: `MergeSubPops`)

Merge subpopulations This operator merges subpopulations `subPops` to a single subpopulation. If `subPops` is ignored, all subpopulations will be merged.

class mergeSubPops(*subPops=[], removeEmptySubPops=False, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
subPops: Subpopulations to be merged. Default to all.

merge subpopulations

apply(*pop*)
Apply a `mergeSubPops` operator

clone()
Deep copy of a `mergeSubPops` operator

2.3.5 Class `resizeSubPops` (Function form: `ResizeSubPops`)

Resize subpopulations This operator resize subpopulations `subPops` to a another size. If `subPops` is ignored, all subpopulations will be resized. If the new size is smaller than the original one, the remaining individuals are discarded. If the new size if greater, individuals will be copied again if propagate is true, and be empty otherwise.

class resizeSubPops(*newSizes=[], subPops=[], propagate=True, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
newSizes: Of the specified (or all) subpopulations.

propagate: If true (default) and the new size if greater than the original size, individuals will be copied over.

subPops: Subpopulations to be resized. Default to all.

resize subpopulations

apply(*pop*)
Apply a `resizeSubPops` operator

clone()
Deep copy of a `resizeSubPops` operator

2.4 Mutation

2.4.1 Class `mutator`

Base class of all mutators. The base class of all functional mutators. It is not supposed to be called directly.

Every mutator can specify `rate` (equal rate or different rates for different loci) and a vector of applicable loci (default to all but should have the same length as `rate` if `rate` has length greater than one).

Maximum allele can be specified as well but more parameters, if needed, should be implemented by individual mutator classes.

There are numbers of possible allelic states. Most theoretical studies assume an infinite number of allelic states to avoid any homoplasy. If it facilitates any analysis, this is however extremely unrealistic.

class `mutator` (*rate=[], loci=[], maxAllele=0, output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

loci: A vector of locus indexes. Will be ignored only when single rate is specified. Default to all loci.

maxAllele: Maximum allowed allele. Interpreted by each sub mutator class. Default to `pop.maxAllele()`.

rate: Can be a number (uniform rate) or an array of mutation rates (the same length as `loci`)

create a mutator, do not call this constructor directly All mutators have the following common parameters. However, the actual meaning of these parameters may vary according to different models. The only differences between the following mutators are the way they actually mutate an allele, and corresponding input parameters. The number of mutation events at each locus is recorded and can be accessed from the `mutationCount` or `mutationCounts` functions.

`apply` (*pop*)
Apply a mutator

`clone` ()
Deep copy of a `mutator`

`maxAllele` ()
Return maximum allowable allele number

`mutate` (*allele*)
Describe how to mutate a single allele

`mutationCount` (*locus*)
Return mutation count at `locus`

`mutationCounts` ()
Return mutation counts

`rate` ()
Return the mutation rate

`setMaxAllele` (*maxAllele*)
Set maximum allowable allele

`setRate` (*rate, loci=[]*)
Set an array of mutation rates

2.4.2 Class `kamMutator` (Function form: `KamMutate`)

K-Allele Model mutator. This mutator mutate an allele to another allelic state with equal probability. The specified mutation rate is actually the 'probability to mutate'. So the mutation rate to any other allelic state is actually $\frac{rate}{K-1}$, where K is specified by parameter `maxAllele`.

class `kamMutator` (*rate=[], loci=[], maxAllele=0, output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

maxAllele: Maximum allele that can be mutated to. For binary libraries, allelic states will be `[0, maxAllele]`. Otherwise, they are `[1, maxAllele]`.

rate: Mutation rate. It is the 'probability to mutate'. The actual mutation rate to any of the other `K-1` allelic states are `rate/(K-1)`.

create a K-Allele Model mutator Please see class `mutator` for the descriptions of other parameters.

clone()
Deep copy of a kamMutator

mutate(*allele*)
Mutate to a state other than current state with equal probability

2.4.3 Class smmMutator (Function form: SmmMutate)

The stepwise mutation model. The *Stepwise Mutation Model* (SMM) assumes that alleles are represented by integer values and that a mutation either increases or decreases the allele value by one. For variable number tandem repeats(VNTR) loci, the allele value is generally taken as the number of tandem repeats in the DNA sequence.

class smmMutator(*rate=[], loci=[], maxAllele=0, incProb=0.5, output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
incProb: Probability to increase allele state. Default to 0.5.

create a SMM mutator The SMM is developed for allozymes. It provides better description for these kinds of evolutionary processes. Please see class mutator for the descriptions of other parameters.

clone()
Deep copy of a smmMutator

2.4.4 Class gsmMutator (Function form: GsmMutate)

Generalized stepwise mutation model The *Generalized Stepwise Mutation model* (GSM) is an extension to the stepwise mutation model. This model assumes that alleles are represented by integer values and that a mutation either increases or decreases the allele value by a random value. In other words, in this model the change in the allelic state is drawn from a random distribution. A *geometric generalized stepwise model* uses a geometric distribution with parameter p, which has mean $\frac{p}{1-p}$ and variance $\frac{p}{(1-p)^2}$.

gsmMutator implements both models. If you specify a Python function without a parameter, this mutator will use its return value each time a mutation occur; otherwise, a parameter p should be provided and the mutator will act as a geometric generalized stepwise model.

class gsmMutator(*rate=[], loci=[], maxAllele=0, incProb=0.5, p=0, func=None, output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
func: A function that returns the number of steps. This function does not accept any parameter.

incProb: Probability to increase allele state. Default to 0.5.

create a gsmMutatorThe GSM model is developed for allozymes. It provides better description for these kinds of evolutionary processes. Please see class mutator for the descriptions of other parameters.

clone()
Deep copy of a gsmMutator

mutate(*allele*)
Mutate according to the GSM model

2.4.5 Class pyMutator (Function form: PyMutate)

A hybrid mutator. Parameters such as mutation rate of this operator are set just like others and you are supposed to provide a Python function to return a new allele state given an old state. pyMutator will choose an allele as usual and call your function to mutate it to another allele.

class pyMutator(*rate=[], loci=[], maxAllele=0, func=None, output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
create a pyMutator

clone()
Deep copy of a pyMutator

mutate(*allele*)
Mutate according to the mixed model

2.4.6 Class pointMutator (Function form: PointMutate)

Point mutator Mutate specified individuals at specified loci to a spcified allele. I.e., this is a non-random mutator used to introduce diseases etc. pointMutator, as its name suggest, does point mutation. This mutator will turn alleles at loci on the first chromosome copy to toAllele for individual inds. You can specify atPloidy to mutate other, or all ploidy copies.

class pointMutator(*loci, toAllele, atPloidy=[], inds=[], output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
inds: Individuals who will mutate

toAllele: Allele that will be mutate to

create a pointMutator Please see class mutator for the descriptions of other parameters.

apply(*pop*)
Apply a pointMutator

clone()
Deep copy of a pointMutator

mutationCount(*locus*)
Return mutation count at locus

mutationCounts()
Return mutation counts

2.5 Recombination and gene conversion

2.5.1 Class recombinator

Recombination and conversion In simuPOP, only one recombinator is provided. Recombination events between loci a/b and b/c are independent, otherwise there will be some linkage between loci. Users need to specify physical recombination rate between adjacent loci. In addition, for the recombinator

- it only works for diploid (and for females in haplodiploid) populations.
- the recombination rate must be comprised between 0.0 and 0.5. A recombination rate of 0.0 means that the loci are completely linked, and thus behave together as a single linked locus. A recombination rate of 0.5 is equivalent to free of recombination. All other values between 0.0 and 0.5 will represent various linkage intensities between adjacent pairs of loci. The recombination rate is equivalent to 1-linkage and represents the probability that the allele at the next locus is randomly drawn.
- it works for selfing. I.e., when only one parent is provided, it will be recombined twice, producing both maternal and paternal chromosomes of the offspring.
- conversion is allowed. Note that conversion will nullify many recombination events, depending on the parameters chosen.

class `recombinator`(*intensity=-1, rate=[], afterLoci=[], maleIntensity=-1, maleRate=[], maleAfterLoci=[], convProb=0, convMode=CONVERT_NumMarkers, convParam=1., begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

afterLoci: An array of locus indexes. Recombination will occur after these loci. If `rate` is also specified, they should have the same length. Default to all loci (but meaningless for those loci located at the end of a chromosome). If this parameter is given, it should be ordered, and can not include loci at the end of a chromosome.

convMode: Conversion mode, determines how track length is determined.

- CONVERT_NumMarkers Converts a fixed number of markers.
- CONVERT_GeometricDistribution An geometric distribution is used to determine how many markers will be converted.
- CONVERT_TractLength Converts a fixed length of tract.
- CONVERT_ExponentialDistribution An exponential distribution with parameter `convLen` will be used to determine track length.

convParam: Parameter for the conversion process. The exact meaning of this parameter is determined by `convMode`. Note that

- conversion tract length is usually short, and is estimated to be between 337 and 456 bp, with overall range between maybe 50 - 2500 bp.
- simuPOP does not impose a unit for marker distance so your choice of `convParam` needs to be consistent with your unit. In the HapMap dataset, cM is usually assumed and marker distances are around 10kb (0.001cM ≈ 1kb). Gene conversion can largely be ignored. This is important when you use distance based conversion mode such as `CONVERT_TrackLength` or `CONVERT_ExponentialDistribution`.
- After a track length is determined, if a second recombination event happens within this region, the track length will be shortened. Note that conversion is identical to double recombination under this context.

convProb: The probability of conversion event among all recombination events. When a recombination event happens, it may become a recombination event if the Holliday junction is resolved/repaired successfully, or a conversion event if the junction is not resolved/repaired. The default `convProb` is 0, meaning no conversion event at all. Note that the ratio of conversion to recombination events varies greatly from study to study, ranging from 0.1 to 15 (Chen et al, Nature Review Genetics, 2007). This translate to 0.1/0.9$\tilde{0}$.1 to 15/16$\tilde{0}$.94 of this parameter. When `convProb` is 1, all recombination events will be conversion events.

haplodiploid: If set to true, the first copy of paternal chromosomes is copied directly as the paternal chromosomes of the offspring. This is because haplodiploid male has only one set of chromosome.

intensity: Intensity of recombination. The actual recombination rate between two loci is determined by `intensity*locus distance (between them)`.

maleAfterLoci: If given, males will recombine at different locations.

maleIntensity: Recombination intensity for male individuals. If given, parameter `intensity` will be considered as female intensity.

maleRate: Recombination rate for male individuals. If given, parameter `rate` will be considered as female recombination rate.

rate: Recombination rate regardless of locus distance after all `afterLoci`. It can also be an array of recombination rates. Should have the same length as `afterLoci` or `totNumOfLoci()`. The recombination rates are independent of locus distance.

recombine chromosomes from parents

Note There is no recombination between sex chromosomes of male individuals if `sexChrom()=True`. This may change later if the exchanges of genes between pseudoautosomal regions of `XY` need to be modeled.

`clone`()

Deep copy of a recombinator

convCount(*size*)
Return the count of conversion of a certain size (only valid in standard modules)

convCounts()
Return the count of conversions of all sizes (only valid in standard modules)

recCount(*locus*)
Return recombination count at a locus (only valid in standard modules)

recCounts()
Return recombination counts (only valid in standard modules)

2.6 Selection

2.6.1 Class `selector`

A base selection operator for all selectors. Genetic selection is tricky to simulate since there are many different *fitness* values and many different ways to apply selection. simuPOP employs an *'ability-to-mate'* approach. Namely, the probability that an individual will be chosen for mating is proportional to its fitness value. More specifically,

- `PreMating` selectors assign fitness values to each individual, and mark part or all subpopulations as under selection.
- during sexless mating (e.g. `binomialSelection` mating scheme), individuals are chosen at probabilities that are proportional to their fitness values. If there are N individuals with fitness values $f_i, i = 1, ..., N$, individual i will have probability $\frac{f_i}{\sum_j f_j}$ to be chosen and passed to the next generation.
- during `randomMating`, males and females are separated. They are chosen from their respective groups in the same manner as `binomialSelection` and mate.

All of the selection operators, when applied, will set an information field `fitness` (configurable) and then mark part or all subpopulations as under selection. (You can use different selectors to simulate various selection intensities for different subpopulations). Then, a *'selector-aware'* mating scheme can select individuals according to their `fitness` information fields. This implies that

- only mating schemes can actually select individuals.
- a selector has to be a `PreMating` operator. This is not a problem when you use the operator form of the selector since its default stage is `PreMating`. However, if you use the function form of the selector in a `pyOperator`, make sure to set the stage of `pyOperator` to `PreMating`.

Note:

You can not apply two selectors to the same subpopulation, because only one fitness value is allowed for each individual.

class selector(*subPops=[], stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["fitness"]*)
subPop: A shortcut to `subPops=[subPop]`

subPops: Subpopulations that the selector will apply to. Default to all.

create a selector

apply(*pop*)
Set fitness to all individuals. No selection will happen!

clone()
Deep copy of a selector

2.6.2 Class `mapSelector` (Function form: `MapSelector`, Applicable to `all ploidy`)

Selection according to the genotype at one or more loci This map selector implements selection according to genotype at one or more loci. A user provided dictionary (map) of genotypes will be used in this selector to set each individual's fitness value.

class `mapSelector` (*loci, fitness, phase=False, subPops=[], stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["fitness"]*)

fitness: A dictionary of fitness values. The genotype must be in the form of `'a-b'` for a single locus, and `'a-b|c-d|e-f'` for multi-loci. In the haploid case, the genotype should be specified in the form of `'a'` for single locus, and `'a|b|c'` for multi-locus models.

loci: The locus indexes. The genotypes at these loci will be used to determine the fitness value.

locus: The locus index. A shortcut to `loci=[locus]`

output: And other parameters please refer to help (`baseOperator.__init__`)

phase: If `True`, genotypes `a-b` and `b-a` will have different fitness values. Default to `False`.

create a map selector

`clone`()

Deep copy of a map selector

`indFitness`(*ind, gen*)

Calculate/return the fitness value, currently assuming diploid

2.6.3 Class `maSelector` (Function form: `MaSelect`)

Multiple allele selector (selection according to wildtype or diseased alleles) This is called 'multiple-allele' selector. It separates alleles into two groups: wildtype and diseased alleles. Wildtype alleles are specified by parameter `wildtype` and any other alleles are considered as diseased alleles.This selector accepts an array of fitness values:

- For single-locus, `fitness` is the fitness for genotypes AA, Aa, aa, while A stands for wildtype alleles.
- For a two-locus model, `fitness` is the fitness for genotypes AABB, AABb, AAbb, AaBB, AbBb, Aabb, aaBB, aaBb and aaBb.
- For a model with more than two loci, use a table of length 3^n in a order similar to the two-locus model.

class `maSelector` (*loci, fitness, wildtype, subPops=[], stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["fitness"]*)

fitness: For the single locus case, `fitness` is an array of fitness of AA, Aa, aa. A is the wildtype group. In the case of multiple loci, fitness should be in the order of AABB, AABb, AAbb, AaBB, AaBb, Aabb, aaBB, aaBb, aabb.

output: And other parameters please refer to help (`baseOperator.__init__`)

wildtype: An array of alleles in the wildtype group. Any other alleles are considered to be diseased alleles. Default to `[0]`.

create a multiple allele selector Please refer to `baseOperator` for other parameter descriptions.

Note

- `maSelector` only works for diploid populations.
- `wildtype` alleles at all loci are the same.

`clone`()

Deep copy of a `maSelector`

`indFitness`(*ind, gen*)

Calculate/return the fitness value, currently assuming diploid

2.6.4 Class `mlSelector` (Function form: `MlSelect`)

Selection according to genotypes at multiple loci in a multiplicative model This selector is a 'multiple-locus model' selector. The selector takes a vector of selectors (can not be another `mlSelector`) and evaluate the fitness of an individual as the product or sum of individual fitness values. The mode is determined by parameter `mode`, which takes one of the following values

- `SEL_Multiplicative`: the fitness is calculated as $f = \prod_i f_i$, where f_i is the single-locus fitness value.
- `SEL_Additive`: the fitness is calculated as $f = \max(0, 1 - \sum_i (1 - f_i))$. f will be set to `0` when $f < 0$.

class `mlSelector`(*selectors, mode=SEL_Multiplicative, subPops=[], stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["fitness"]*)

selectors: A list of selectors

create a multiple-locus selector Please refer to `mapSelector` for other parameter descriptions.

`clone`()

Deep copy of a `mlSelector`

`indFitness`(*ind, gen*)

Calculate/return the fitness value, currently assuming diploid

2.6.5 Class `pySelector` (Function form: `PySelect`)

Selection using user provided function This selector assigns fitness values by calling a user provided function. It accepts a list of loci and a Python function `func`. For each individual, this operator will pass the genotypes at these loci, generation number, and optionally values at some information fields to this function. The return value is treated as the fitness value. The genotypes are arranged in the order of `0-0,0-1,1-0,1-1` etc. where X-Y represents locus X - ploidy Y. More specifically, `func` can be

- `func(geno, gen)` if `infoFields` has length 0 or 1.
- `func(geno, gen, fields)` when `infoFields` has more than 1 fields. Values of fields 1, 2, ... will be passed. Both `geno` and `fields` should be a list.

class `pySelector`(*loci, func, subPops=[], stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["fitness"]*)

func: A Python function that accepts genotypes at specified loci, generation number, and optionally information fields. It returns the fitness value.

infoFields: If specified, the first field should be the information field to save calculated fitness value (should be 'fitness' in most cases). The values of the rest of the information fields (if available) will also be passed to the user defined penetrance function.

loci: Susceptibility loci. The genotype at these loci will be passed to `func`.

output: And other parameters please refer to help (`baseOperator.__init__`)

create a Python hybrid selector

`clone`()

Deep copy of a `pySelector`

`indFitness`(*ind, gen*)

Calculate/return the fitness value, currently assuming diploid

2.7 Penetrance

2.7.1 Class `penetrance`

Base class of all penetrance operators. Penetrance is the probability that one will have the disease when he has certain genotype(s). An individual will be randomly marked as affected/unaffected according to his/her penetrance value. For example, an individual will have probability 0.8 to be affected if the penetrance is 0.8.

Penetrance can be applied at any stage (default to `DuringMating`). When a penetrance operator is applied, it calculates the penetrance value of each offspring and assigns affected status accordingly. Penetrance can also be used `PreMating` or `PostMating`. In these cases, the affected status will be set to all individuals according to their penetrance values.

Penetrance values are usually not saved. If you would like to know the penetrance value, you need to

- use `addInfoField('penetrance')` to the population to analyze. (Or use `infoFields` parameter of the population constructor), and
- use e.g., `mlPenetrance(...., infoFields=['penetrance'])` to add the penetrance field to the penetrance operator you use. You may choose a name other than `'penetrance'` as long as the field names for the operator and population match.

Penetrance functions can be applied to the current, all, or certain number of ancestral generations. This is controlled by the `ancestralGen` parameter, which is default to `-1` (all available ancestral generations). You can set it to `0` if you only need affection status for the current generation, or specify a number `n` for the number of ancestral generations (n + 1 total generations) to process. Note that the `ancestralGen` parameter is ignored if the penetrance operator is used as a during mating operator.

class penetrance(*ancestralGen=-1, stage=DuringMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

ancestralGen: If this parameter is set to be `0`, apply penetrance to the current generation; if `-1`, apply to all generations; otherwise, apply to the specified numbers of ancestral generations.

infoFields: If one field is specified, it will be used to store penetrance values.

stage: Specify the stage this operator will be applied. Default to `DuringMating`.

create a penetrance operator

apply(*pop*)

Set penetrance to all individuals and record penetrance if requested

clone()

Deep copy of a penetrance operator

penet()

Calculate/return penetrance etc.

2.7.2 Class `mapPenetrance` (Function form: `MapPenetrance`)

Penetrance according to the genotype at one locus Assign penetrance using a table with keys 'X-Y' where X and Y are allele numbers.

class mapPenetrance(*loci, penet, phase=False, ancestralGen=-1, stage=DuringMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

loci: The locus indexes. The genotypes of these loci will be used to determine penetrance.

locus: The locus index. Shortcut to `loci=[locus]`

output: And other parameters please refer to help(baseOperator.__init__)

penet: A dictionary of penetrance. The genotype must be in the form of 'a-b' for a single locus.

phase: If `True`, `a/b` and `b/a` will have different penetrance values. Default to `False`.

create a map penetrance operator

clone()
Deep copy of a map penetrance operator

2.7.3 Class `maPenetrance` (Function form: `MaPenetrance`)

Multiple allele penetrance operator This is called 'multiple-allele' penetrance. It separates alleles into two groups: wildtype and diseased alleles. Wildtype alleles are specified by parameter `wildtype` and any other alleles are considered as diseased alleles. `maPenetrance` accepts an array of penetrance for AA, Aa, aa in the single-locus case, and a longer table for the multi-locus case. Penetrance is then set for any given genotype.

class maPenetrance(*loci, penet, wildtype, ancestralGen=-1, stage=DuringMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
loci: The locus indexes. The genotypes of these loci will be examed.

locus: The locus index. The genotype of this locus will be used to determine penetrance.

output: And other parameters please refer to help(baseOperator.__init__)

penet: An array of penetrance values of AA, Aa, aa. A is the wild type group. In the case of multiple loci, penetrance should be in the order of AABB, AABb, AAbb, AaBB, AaBb, Aabb, aaBB, aaBb, aabb.

wildtype: An array of alleles in the wildtype group. Any other alleles will be considered as in the diseased allele group.

create a multiple allele penetrance operator (penetrance according to diseased or wildtype alleles)

clone()
Deep copy of a multi-allele penetrance operator

penet(*ind*)
Currently assuming diploid

2.7.4 Class `mlPenetrance` (Function form: `MlPenetrance`)

Penetrance according to the genotype according to a multiple loci multiplicative model This is the 'multiple-locus' penetrnace calculator. It accepts a list of penetrances and combine them according to the `mode` parameter, which takes one of the following values:

- `PEN_Multiplicative`: the penetrance is calculated as $f = \prod f_i$.
- `PEN_Additive`: the penetrance is calculated as $f = \min(1, \sum f_i)$. f will be set to `1` when $f < 0$. In this case, s_i are added, not f_i directly.
- `PEN_Heterogeneity`: the penetrance is calculated as $f = 1 - \prod (1 - f_i)$.

Please refer to Neil Risch (1990) for detailed information about these models.

class mlPenetrance(*peneOps, mode=PEN_Multiplicative, ancestralGen=-1, stage=DuringMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
mode: Can be one of `PEN_Multiplicative`, `PEN_Additive`, and `PEN_Heterogeneity`

peneOps: A list of penetrance operators

create a multiple locus penetrance operator

clone()
Deep copy of a multi-loci penetrance operator

penet(*ind*)
Currently assuming diploid

2.7.5 Class `pyPenetrance` (Function form: `PyPenetrance`)

Assign penetrance values by calling a user provided function For each individual, the penetrance is determined by a user-defined penetrance function `func`. This function takes genetypes at specified loci, and optionally values of specified information fields. The return value is considered as the penetrance for this individual.More specifically, `func` can be

- `func(geno)` if `infoFields` has length 0 or 1.
- `func(geno, fields)` when `infoFields` has more than 1 fields. Both parameters should be an list.

class pyPenetrance(*loci, func, ancestralGen=-1, stage=DuringMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
func: A user-defined Python function that accepts an array of genotypes at specified loci and return a penetrance value. The return value should be between `0` and `1`.

infoFields: If specified, the first field should be the information field to save calculated penetrance value. The values of the rest of the information fields (if available) will also be passed to the user defined penetrance function.

loci: The genotypes at these loci will be passed to the provided Python function in the form of `loc1_1, loc1_2, loc2_1, loc2_2, ...` if the individuals are diploid.

output: And other parameters please refer to help (`baseOperator.__init__`)

provide locus and penetrance for 11, 12, 13 (in the form of dictionary)

clone()
Deep copy of a Python penetrance operator

penet(*ind*)
Currently assuming diploid

2.8 Quantitative Trait

2.8.1 Class `quanTrait`

Base class of quantitative trait Quantitative trait is the measure of certain phenotype for given genotype. Quantitative trait is similar to penetrance in that the consequence of penetrance is binary: affected or unaffected; while it is continuous for quantitative trait.

In simuPOP, different operators or functions were implemented to calculate quantitative traits for each individual and store the values in the information fields specified by the user (default to `qtrait`). The quantitative trait operators also accept the `ancestralGen` parameter to control the number of generations for which the `qtrait` information field will be set.

class quanTrait(*ancestralGen=-1, stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["qtrait"]*)
create a quantitative trait operator

apply(*pop*)
Set `qtrait` to all individual

clone()
Deep copy of a quantitative trait operator

qtrait()
Calculate/return quantitative trait etc.

2.8.2 Class `mapQuanTrait` (Function form: `MapQuanTrait`)

Quantitative trait according to genotype at one locus Assign quantitative trait using a table with keys 'X-Y' where X and Y are allele numbers. If parameter `sigma` is not zero, the return value is the sum of the trait plus $N\left(0,\sigma^2\right)$. This random part is usually considered as the environmental factor of the trait.

class mapQuanTrait(*loci, qtrait, sigma=0, phase=False, ancestralGen=-1, stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["qtrait"]*)
loci: An array of locus indexes. The quantitative trait is determined by genotypes at these loci.

locus: The locus index. The quantitative trait is determined by genotype at this locus.

output: And other parameters please refer to help (`baseOperator.__init__`)

phase: If `True`, a/b and b/a will have different quantitative trait values. Default to `False`.

qtrait: A dictionary of quantitative traits. The genotype must be in the form of 'a-b'. This is the mean of the quantitative trait. The actual trait value will be $N\left(mean,\sigma^2\right)$. For multiple loci, the form is 'a-blc-dle-f' etc.

sigma: Standard deviation of the environmental factor $N\left(0,\sigma^2\right)$.

create a map quantitative trait operator

clone()
Deep copy of a map quantitative trait operator

qtrait(*ind*)
Currently assuming diploid

2.8.3 Class `maQuanTrait` (Function form: `MaQuanTrait`)

Multiple allele quantitative trait (quantitative trait according to disease or wildtype alleles) This is called 'multiple-allele' quantitative trait. It separates alleles into two groups: wildtype and diseased alleles. Wildtype alleles are specified by parameter `wildtype` and any other alleles are considered as diseased alleles. `maQuanTrait` accepts an array of fitness. Quantitative trait is then set for any given genotype. A standard normal distribution $N\left(0,\sigma^2\right)$ will be added to the returned trait value.

class maQuanTrait(*loci, qtrait, wildtype, sigma=[], ancestralGen=-1, stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["qtrait"]*)
output: And other parameters please refer to help(`baseOperator.__init__`)

qtrait: An array of quantitative traits of AA, Aa, aa. A is the wildtype group

sigma: An array of standard deviations for each of the trait genotype (AA, Aa, aa)

wildtype: An array of alleles in the wildtype group. Any other alleles will be considered as diseased alleles. Default to `[0]`.

create a multiple allele quantitative trait operator Please refer to `quanTrait` for other parameter descriptions.

clone()
Deep copy of a multiple allele quantitative trait

qtrait(*ind*)
Currently assuming diploid

2.8.4 Class `mlQuanTrait` (Function form: `MlQuanTrait`)

Quantitative trait according to genotypes from a multiple loci multiplicative model Operator `mlQuanTrait` is a 'multiple-locus' quantitative trait calculator. It accepts a list of quantitative traits and combine them according to the `mode` parameter, which takes one of the following values

- `QT_Multiplicative`: the mean of the quantitative trait is calculated as $f = \prod f_i$.
- `QT_Additive`: the mean of the quantitative trait is calculated as $f = \sum f_i$.

Note that all σ_i (for f_i) and σ (for f) will be considered. I.e, the trait value should be

$$f = \sum_i \left(f_i + N\left(0, \sigma_i^2\right)\right) + \sigma^2$$

for `QT_Additive` case. If this is not desired, you can set some of the σ to zero.

class `mlQuanTrait` (*qtraits, mode=QT_Multiplicative, sigma=0, ancestralGen=-1, stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["qtrait"]*)

mode: Can be one of `QT_Multiplicative` and `QT_Additive`

qtraits: A list of quantitative traits

create a multiple locus quantitative trait operator Please refer to `quanTrait` for other parameter descriptions.

`clone`()

Deep copy of a multiple loci quantitative trait operator

`qtrait`(*ind*)

Currently assuming diploid

2.8.5 Class `pyQuanTrait` (Function form: `PyQuanTrait`)

Quantitative trait using a user provided function For each individual, a user provided function is used to calculate quantitative trait.

class `pyQuanTrait` (*loci, func, ancestralGen=-1, stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["qtrait"]*)

func: A Python function that accepts genotypes at specified loci and returns the quantitative trait value.

loci: The genotypes at these loci will be passed to `func`.

output: And other parameters please refer to help(`baseOperator.__init__`)

create a Python quantitative trait operator Please refer to `quanTrait` for other parameter descriptions.

`clone`()

Deep copy of a Python quantitative trait operator

`qtrait`(*ind*)

Currently assuming diploid

2.9 Ascertainment

2.9.1 Class `sample`

Base class of other sample operator Ascertainment/sampling refers to the ways of selecting individuals from a population. In simuPOP, ascerntainment operators create sample populations that can be accessed from the population's

local namespace. All the ascertainment operators work like this except for `pySubset` which shrink the population itself.

Individuals in sampled populations may or may not keep their original order but their indexes in the whole population are stored in an information field `oldindex`. This is to say, you can use `ind.info('oldindex')` to check the original position of an individual.

Two forms of sample size specification are supported: with or without subpopulation structure. For example, the `size` parameter of `randomSample` can be a number or an array (which has the length of the number of subpopulations). If a number is given, a sample will be drawn from the whole population, regardless of the population structure. If an array is given, individuals will be drawn from each subpopulation `sp` according to `size[sp]`.

An important special case of sample size specification occurs when `size=[]` (default). In this case, usually all qualified individuals will be returned.

The function forms of these operators are a little different from others. They do return a value: an array of samples.

class sample(*name="sample", nameExpr="", times=1, saveAs="", saveAsExpr="", format="auto", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

format: Format to save the samples

name: Name of the sample in the local namespace. This variable is an array of populations of size `times`. Default to `sample`.

nameExpr: Expression version of parameter `name`. If both `name` and `nameExpr` are empty, sample populations will not be saved in the population's local namespace. This expression will be evaluated dynamically in population's local namespace.

saveAs: Filename to save the samples

saveAsExpr: Expression version of parameter `saveAs`. It will be evaluated dynamically in population's local namespace.

times: How many times to sample from the population. This is usually `1`, but we may want to take several random samples.

draw a sample Please refer to `baseOperator::__init__` for other parameter descriptions.

apply(*pop*)
Apply the `sample` operator

clone()
Deep copy of a `sample` operator

samples(*pop*)
Return the samples

2.9.2 Class `pySubset` (Function form: `PySubset`)

Shrink population This operator shrinks a population according to a given array or the `subPopID()` value of each individual. Individuals with negative subpopulation IDs will be removed.

class pySubset(*keep=[], stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

keep: An array of individual subpopulation IDs

create a `pySubset` operator

apply(*pop*)
Apply the `pySubset` operator

clone()
Deep copy of a `pySubset` operator

2.9.3 Class `pySample` (Function form: `PySample`)

Python sampler. A Python sampler that generate a sample with given individuals. This sampler accepts a Python array with elements that will be assigned to individuals as their subpopulation IDs. Individuals with positive subpopulation IDs will then be picked out and form a sample.

class `pySample`(*keep, keepAncestralPops=-1, name="sample", nameExpr="", times=1, saveAs="", saveAsExpr="", format="auto", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

keep: Subpopulation IDs of all individuals

keepAncestralPop: The number of ancestral populations that will be kept. If `-1` is given, keep all ancestral populations (default). If `0` is given, no ancestral population will be kept.

create a Python sampler Please refer to class `sample` for other parameter descriptions.

`clone`()

Deep copy of a Python sampler

`drawsample`(*pop*)

Draw a Python sample

2.9.4 Class `randomSample` (Function form: `RandomSample`)

Randomly draw a sample from a population This operator will randomly choose `size` individuals (or `size[i]` individuals from subpopulation `i`) and return a new population. The function form of this operator returns the samples directly. This operator keeps samples in an array `name` in the local namespace. You may access them through `dvars()` or `vars()` functions.

The original subpopulation structure or boundary is kept in the samples.

class `randomSample`(*size=[], name="sample", nameExpr="", times=1, saveAs="", saveAsExpr="", format="auto", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

size: Size of the sample. It can be either a number which represents the overall sample size, regardless of the population structure; or an array which represents the number of individuals drawn from each subpopulation.

draw a random sample, regardless of the affectedness status Please refer to class `sample` for other parameter descriptions.

Note Ancestral populations will not be copied to the samples.

`clone`()

Deep copy of a `randomSample` operator

2.9.5 Class `caseControlSample` (Function form: `CaseControlSample`)

Draw a case-control sample from a population This operator will randomly choose `cases` affected individuals and `controls` unaffected individuals as a sample. The affectedness status is usually set by penetrance functions or operators. The sample populations will have two subpopulations: cases and controls.

You may specify the number of cases and the number of controls from each subpopulation using the array form of the parameters. The sample population will still have only two subpoulations (cases and controls) though.

A special case of this sampling scheme occurs when one of or both `cases` and `controls` are omitted (zeros). In this case, all cases and/or controls are chosen. If both parameters are omitted, the sample is effectively the same population with affected and unaffected individuals separated into two subpopulations.

class caseControlSample(*cases=[], controls=[], spSample=False, name="sample", nameExpr="", times=1, saveAs="", saveAsExpr="", format="auto", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

cases: The number of cases, or an array of the numbers of cases from each subpopulation

controls: The number of controls, or an array of the numbers of controls from each subpopulation

draw cases and controls as a sample Please refer to class `sample` for other parameter descriptions.

clone()

Deep copy of a `caseControlSample` operator

2.9.6 Class `affectedSibpairSample` (Function form: `AffectedSibpairSample`)

Draw an affected sibling pair sample Special preparation for the population is needed in order to use this operator. Obviously, to obtain affected sibling pairs, we need to know the parents and the affectedness status of each individual. Furthermore, to get parental genotypes, the population should have `ancestralDepth` at least `1`. The most important problem, however, comes from the mating scheme we are using.

`randomMating()` is usually used for diploid populations. The *real random* mating requires that a mating will generate only one offspring. Since parents are chosen with replacement, a parent can have multiple offspring with different parents. On the other hand, it is very unlikely that two offspring will have the same parents. The probability of having a sibling for an offspring is $\frac{1}{N^2}$ (if do not consider selection). Therefore, we will have to allow multiple offspring per mating at the cost of small effective population size.

class affectedSibpairSample(*size=[], chooseUnaffected=False, countOnly=False, name="sample", nameExpr="", times=1, saveAs="", saveAsExpr="", format="auto", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["father_idx", "mother_idx"]*)

chooseUnaffected: Instead of affected sibpairs, choose unaffected families.

countOnly: Set variables about the number of affected sibpairs, do not actually draw the sample

size: The number of affected sibling pairs to be sampled. Can be a number or an array. If a number is given, it is the total number of sibpairs, ignoring the population structure. Otherwise, specified numbers of sibpairs are sampled from subpopulations. If `size` is unspecified, this operator will return all affected sibpairs.

draw an affected sibling pair sample Please refer to class `sample` for other parameter descriptions.

clone()

Deep copy of a `affectedSibpairSample` operator

drawsample(*pop*)

Draw a sample

prepareSample(*pop*)

Preparation before drawing a sample

2.9.7 Class `largePedigreeSample`

Draw a large pedigree sample

class largePedigreeSample(*size=[], minTotalSize=0, maxOffspring=5, minPedSize=5, minAffected=0, countOnly=False, name="sample", nameExpr="", times=1, saveAs="", saveAsExpr="", format="auto", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["father_idx", "mother_idx"]*)

countOnly: Set variables about the number of affected sibpairs, do not actually draw the sample.

maxOffspring: The maximum number of offspring a parent may have

minAffected: The minimal number of affected individuals in each pedigree. Default to `0`.

minPedSize: The minimal pedigree size. Default to `5`.

minTotalSize: The minimum number of individuals in the sample

draw a large pedigree sample Please refer to class `sample` for other parameter descriptions.

`clone``()`
Deep copy of a `largePedigreeSample` operator

`drawsample``(`*pop*`)`
Draw a a large pedigree sample

`prepareSample``(`*pop*`)`
Preparation before drawing a sample

2.9.8 Class `nuclearFamilySample`

Draw a nuclear family sample

class `nuclearFamilySample``(`*size=[], minTotalSize=0, maxOffspring=5, minPedSize=5, minAffected=0, countOnly=False, name="sample", nameExpr="", times=1, saveAs="", saveAsExpr="", format="auto", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=["father_idx", "mother_idx"]*`)`
draw a nuclear family sample Please refer to class `sample` for parameter descriptions.

`clone``()`
Deep copy of a `nuclearFamilySample` operator

`drawsample``(`*pop*`)`
Draw a nuclear family sample

`prepareSample``(`*pop*`)`
Preparation before drawing a sample

2.10 Statistics Calculation

2.10.1 Class `stator`

Base class of all the statistics calculator Operator `stator` calculates various basic statistics for the population and set variables in the local namespace. Other operators or functions can refer to the results from the namespace after `stat` is applied.

class `stator``(`*output="", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*`)`
create a stator

`clone``()`
Deep copy of a stator

2.10.2 Class `stat` (Function form: `Stat`)

Calculate statistics Operator `stat` calculates various basic statistics for the population and sets variables in the local namespace. Other operators or functions can refer to the results from the namespace after `stat` is applied. `Stat` is the function form of the operator.

Note that these statistics are dependent to each other. For example, heterotype and allele frequencies of related loci will be automatically calculated if linkage diseqilibrium is requested.

class `stat` (*popSize=False, numOfMale=False, numOfMale_param={}, numOfAffected=False, numOfAffected_param={}, numOfAlleles=[], numOfAlleles_param={}, alleleFreq=[], alleleFreq_param={}, heteroFreq=[], expHetero=[], expHetero_param={}, homoFreq=[], genoFreq=[], genoFreq_param={}, haploFreq=[], LD=[], LD_param={}, association=[], association_param={}, Fst=[], Fst_param={}, relGroups=[], relLoci=[], rel_param={}, relBySubPop=False, relMethod=[], relMinScored=10, hasPhase=False, midValues=False, output="", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

Fst: Calculate F_{st}, F_{is}, F_{it}. For example, `Fst = [0,1,2]` will calculate F_{st}, F_{is}, F_{it} based on alleles at loci `0`, `1`, `2`. The locus-specific values will be used to calculate `AvgFst`, which is an average value over all alleles (Weir & Cockerham, 1984). Terms and values that match Weir & Cockerham are:

- F (F_{IT}) the correlation of genes within individuals (inbreeding);
- θ (F_{ST}) the correlation of genes of difference individuals in the same population (will evaluate for each subpopulation and the whole population)
- f (F_{IS}) the correlation of genes within individuals within populations.

This parameter will set the following variables:

- `Fst[loc]`, `Fis[loc]`, `Fit[loc]`
- `AvgFst`, `AvgFis`, `AvgFit`.

Fst_param: A dictionary of parameters of `Fst` statistics. Can be one or more items choosen from the following options: `Fst`, `Fis`, `Fit`, `AvgFst`, `AvgFis`, and `AvgFit`.

LD: Calculate linkage disequilibria LD, LD' and r^2, given `LD=[ [loc1, loc2], [ loc1, loc2, allele1, allele2], ... ]`. For each item `[loc1, loc2, allele1, allele2]`, D, D' and r^2 will be calculated based on `allele1` at `loc1` and `allele2` at `loc2`. If only two loci are given, the LD values are averaged over all allele pairs. For example, for allele A at locus `1` and allele B at locus `2`,

$$D = P_{AB} - P_A P_B$$

$$D' = D/D_{max}$$

$$D_{max} = \min\left(P_A\left(1-P_B\right), \left(1-P_A\right)P_B\right) \text{ if } D > 0 \min\left(P_A P_B, \left(1-P_A\right)\left(1-P_B\right)\right) \text{ if } D < 0$$

$$r^2 = \frac{D^2}{P_A\left(1-P_A\right)P_B\left(1-P_B\right)}$$

If only one item is specified, the outer `[]` can be ignored. I.e., `LD=[loc1, loc2]` is acceptable. This parameter will set the following variables. Please note that the difference between the data structures used for `ld` and `LD`.

- `ld['loc1-loc2']['allele1-allele2']`, `subPop[sp]['ld']['loc1-loc2']['allele1-allele2']`
- `ld_prime['loc1-loc2']['allele1-allele2']`, `subPop[sp]['ld_prime']['loc1-loc2']['allel`
- `r2['loc1-loc2']['allele1-allele2']`, `subPop[sp]['r2']['loc1-loc2']['allele1-allele2']`
- `LD[loc1][loc2]`, `subPop[sp]['LD'][loc1][loc2]`.
- `LD_prime[loc1][loc2]`, `subPop[sp]['LD_prime'][loc1][loc2]`.
- `R2[loc1][loc2]`, `subPop[sp]['R2'][loc1][loc2]`.

LD_param: A dictionary of parameters of `LD` statistics. Can have key `stat` which is a list of statistics to calculate. Default to all. If any statistics is specified, only those specified will be calculated. For example, you may use `LD_param={LD_prime}` to calculate D' only, where `LD_prime` is a shortcut for `'stat':['LD_prime']`. Other parameters that you may use are:

- `subPop` whether or not calculate statistics for subpopulations.
- `midValues` whether or not keep intermediate results.

alleleFreq: An array of loci at which all allele frequencies will be calculated (`alleleFreq=[loc1, loc2, ...]` where `loc1` etc. are loci where allele frequencies will be calculated). This parameter will set the following variables (`carray` objects); for example, `alleleNum[1][2]` will be the number of allele `2` at locus `1`:

- `alleleNum[a]`, `subPop[sp]['alleleNum'][a]`
- `alleleFreq[a]`, `subPop[sp]['alleleFreq'][a]`.

alleleFreq_param: A dictionary of parameters of `alleleFreq` statistics. Can be one or more items choosen from the following options: `numOfAlleles`, `alleleNum`, and `alleleFreq`.

association: Association measures

association_param: A dictionary of parameters of `association` statistics. Can be one or more items choosen from the following options: `ChiSq`, `ChiSq_P`, `UC_U`, and `CramerV`.

expHetero: An array of loci at which the expected heterozygosities will be calculated (`expHetero=[loc1, loc2, ...]`). The expected heterozygosity is calculated by

$$h_{exp} = 1 - p_i^2,$$

where p_i is the allele frequency of allele i. The following variables will be set:

- `expHetero[loc]`, `subPop[sp]['expHetero'][loc]`.

expHetero_param: A dictionary of parameters of `expHetero` statistics. Can be one or more items choosen from the following options: `subpop` and `midValues`.

genoFreq: An array of loci at which all genotype frequencies will be calculated (`genoFreq=[loc1, loc2, ...]`. You may use parameter `genoFreq_param` to control if `a/b` and `b/a` are the same genotype. This parameter will set the following dictionary variables. Note that unlike list used for `alleleFreq` etc., the indexes `a`, `b` of `genoFreq[loc][a][b]` are dictionary keys, so you will get a *KeyError* when you used a wrong key. You can get around this problem by using expressions like `genoNum[loc].setDefault(a,{})`.

- `genoNum[loc][allele1][allele2]` and `subPop[sp]['genoNum'][loc][allele1][allele2]`, the number of genotype `allele1-allele2` at locus `loc`.
- `genoFreq[loc][allele1][allele2]` and `subPop[sp]['genoFreq'][loc][allele1][allele2]`, the frequency of genotype `allele1-allele2` at locus `loc`.
- genoFreq_param a dictionary of parameters of `phase` = 0 or 1.

haploFreq: A matrix of haplotypes (allele sequences on different loci) to count. For example, `haploFreq = [ [ 0,1,2 ], [1,2] ]` will count all haplotypes on loci 0, 1 and 2; and all haplotypes on loci 1, 2. If only one haplotype is specified, the outer `[]` can be omitted. I.e., `haploFreq=[0,1]` is acceptable. The following dictionary variables will be set with keys `0-1-2` etc. For example, `haploNum['1-2']['5-6']` is the number of allele pair 5, 6 (on loci 1 and 2 respectively) in the population.

- `haploNum[haplo]` and `subPop[sp]['haploNum'][haplo]`, the number of allele sequencies on loci `haplo`.
- `haploFreq[haplo]`, `subPop[sp]['haploFreq'][haplo]`, the frequency of allele sequencies on loci `haplo`.

hasPhase: If a/b and b/a are the same genotype. Default to `False`.

heteroFreq: An array of loci at which observed heterozygosities will be calculated (`heteroFreq=[loc1, loc2, ...]`). For each locus, the number and frequency of allele specific and overall heterozygotes will be calcuated and stored in four population variables. For example, `heteroNum[loc][1]` stores number of heterozygotes at locus `loc`, with respect to allele `1`, which is the number of all genotype `1x`

or `x1` where does not equal to `1`. All other genotypes such as `02` are considered as homozygotes when `heteroFreq[loc][1]` is calculated. The overall number of heterozygotes (`HeteroNum[loc]`) is the number of genotype `xy` if `x` does not equal to `y`.

- `HeteroNum[loc]`, `subPop[sp]['HeteroNum'][loc]`, the overall heterozygote count.
- `HeteroFreq[loc]`, `subPop[sp]['HeteroFreq'][loc]`, the overall heterozygote frequency.
- `heteroNum[loc][allele]`, `subPop[sp]['heteroNum'][loc][allele]`, allele-specific heterozygote counts.
- `heteroFreq[loc][allele]`, `subPop[sp]['heteroFreq'][loc][allele]`, allele-specific heterozygote frequency.

homoFreq: An array of loci to calculate observed homozygosities and expected homozygosities (`homoFreq=[loc1, loc2, ...]`). This parameter will calculate the numbers and frequencies of homozygotes **xx** and set the following variables:

- `homoNum[loc]`, `subPop[sp]['homoNum'][loc]`.
- `homoFreq[loc]`, `subPop[sp]['homoFreq'][loc]`.

midValues: Whether or not post intermediate results. Default to `False`. For example, `Fst` will need to calculate allele frequencise. If `midValues` is set to `True`, allele frequencies will be posted as well. This will be helpful in debugging and sometimes in deriving statistics.

numOfAffected: Whether or not count the numbers or proportions of affected and unaffected individuals. This parameter can set the following variables by user's specification:

- `numOfAffected`, `subPop[sp]['numOfAffected']` the number of affected individuals in the population/subpopulation.
- `numOfUnaffected`, `subPop[sp]['numOfUnAffected']` the number of unaffected individuals in the population/subpopulation.
- `propOfAffected`, `subPop[sp]['propOfAffected']` the proportion of affected individuals in the population/subpopulation.
- `propOfUnaffected`, `subPop[sp]['propOfUnAffected']` the proportion of unaffected individuals in the population/subpopulation.

numOfAffected_param: A dictionary of parameters of `numOfAffected` statistics. Can be one or more items choosen from the following options: `numOfAffected`, `propOfAffected`, `numOfUnaffected`, `propOfUnaffected`.

numOfAlleles: An array of loci at which the numbers of distinct alleles will be counted (`numOfAlleles=[loc1, loc2, ...]` where `loc1` etc. are absolute locus indexes). This is done through the calculation of allele frequencies. Therefore, allele frequencies will also be calculated if this statistics is requested. This parameter will set the following variables (`carray` objects of the numbers of alleles for *all loci*). Unrequested loci will have `0` distinct alleles.

- `numOfAlleles`, `subPop[sp]['numOfAlleles']` the number of distinct alleles at each locus. (Calculated only at requested loci.)

numOfAlleles_param: A dictionary of parameters of `numOfAlleles` statistics. Can be one or more items choosen from the following options: `numOfAffected`, `propOfAffected`, `numOfUnaffected`, `propOfUnaffected`.

numOfMale: Whether or not count the numbers or proportions of males and females. This parameter can set the following variables by user's specification:

- `numOfMale`, `subPop[sp]['numOfMale']` the number of males in the population/subpopulation.
- `numOfFemale`, `subPop[sp]['numOfFemale']` the number of females in the population/subpopulation.

- `propOfMale`, `subPop[sp]['propOfMale']` the proportion of males in the population/subpopulation.
- `propOfFemale`, `subPop[sp]['propOfFemale']` the proportion of females in the population/subpopulation.

numOfMale_param: A dictionary of parameters of `numOfMale` statistics. Can be one or more items choosen from the following options: `numOfMale`, `propOfMale`, `numOfFemale`, and `propOfFemale`.

popSize: Whether or not calculate population and virtual subpopulation sizes. This parameter will set the following variables:

- `numSubPop` the number of subpopulations.
- `subPopSize` an array of subpopulation sizes.
- `virtualSubPopSize` (optional) an array of virtual subpopulation sizes. If a subpopulation does not have any virtual subpopulation, the subpopulation size is returned.
- `popSize`, `subPop[sp]['popSize']` the population/subpopulation size.

relGroups: Calculate pairwise relatedness between groups. Can be in the form of either `[[1,2,3],[5,6,7],[8,9]]` or `[2,3,4]`. The first one specifies groups of individuals, while the second specifies subpopulations. By default, relatedness between subpopulations is calculated.

relLoci: Loci on which relatedness values are calculated

relMethod: Method used to calculate relatedness. Can be either `REL_Queller` or `REL_Lynch`. The relatedness values between two individuals, or two groups of individuals are calculated according to Queller & Goodnight (1989) (`method=REL_Queller`) and Lynch et al. (1999) (`method=REL_Lynch`). The results are pairwise relatedness values, in the form of a matrix. Original group or subpopulation numbers are discarded. There is no subpopulation level relatedness value.

rel_param: A dictionary of parameters of relatedness statistics. Can be one or more items choosen from the following options: `Fst`, `Fis`, `Fit`, `AvgFst`, `AvgFis`, and `AvgFit`.

create an `stat` operator

`apply`(*pop*)
Apply the `stat` operator

`clone`()
Deep copy of a `stat` operator

2.11 Expression and Statements

2.11.1 Class `dumper`

Dump the content of a population.

class `dumper`(*alleleOnly=False, infoOnly=False, ancestralPops=False, dispWidth=1, max=100, chrom=[], loci=[], subPop=[], indRange=[], output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
alleleOnly: Only display allele

ancestralPops: Whether or not display ancestral populations. Default to `False`.

chrom: Chromosome(s) to display

dispWidth: Number of characters to display an allele. Default to `1`.

indRange: Range(s) of individuals to display

infoOnly: Only display genotypic information

loci: Loci to display

max: The maximum number of individuals to display. Default to `100`. This is to avoid careless dump of huge populations.

output: Output file. Default to the standard output.

outputExpr: And other parameters: refer to help(baseOperator.__init__)

subPop: Only display subpopulation(s)

dump a population

alleleOnly()
Only show alleles (not structure, gene information?

apply(*pop*)
Apply to one population. It does not check if the operator is activated.

clone()
Deep copy of an operator

infoOnly()
Only show info

setAlleleOnly(*alleleOnly*)
FIXME: No document

setInfoOnly(*infoOnly*)
FIXME: No document

2.11.2 Class `savePopulation`

Save population to a file

class savePopulation(*output="", outputExpr="", format="", compress=True, stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
compress: Obsolete parameter

format: Obsolete parameter

output: Output filename.

outputExpr: An expression that will be evalulated dynamically to determine file name. Parameter `output` will be ignored if this parameter is given.

save population

apply(*pop*)
Apply to one population. It does not check if the operator is activated.

clone()
Deep copy of an operator

2.11.3 Class `pyOutput`

Output a given string. A common usage is to output a new line for the last replicate.

class pyOutput(*str="", output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
str: String to be outputted

Create a `pyOutput` operator that outputs a given string.

apply(*pop*)
Simply output some info

clone()
Deep copy of an operator

setString(*str*)
Set output string.

2.11.4 Class pyEval (Function form: PyEval)

Evaluate an expression Python expressions/statements will be executed when pyEval is applied to a population by using parameters expr/stmts. Statements can also been executed when pyEval is created and destroyed or before expr is executed. The corresponding parameters are preStmts, postStmts and stmts. For example, operator varPlotter uses this feature to initialize R plots and save plots to a file when finished.

class pyEval(*expr="", stmts="", preStmts="", postStmts="", exposePop=False, name="", output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
exposePop: If True, expose the current population as a variable named pop

expr: The expression to be evaluated. The result will be sent to output.

name: Used to let pure Python operator to identify themselves

output: Default to >. I.e., output to standard output.

postStmts: The statement that will be executed when the operator is destroyed

preStmts: The statement that will be executed when the operator is constructed

stmts: The statement that will be executed before the expression

evaluate expressions/statments in the local namespace of a replicate

apply(*pop*)
Apply the pyEval operator

clone()
Deep copy of a pyEval operator

name()
Return the name of an expression The name of a pyEval operator is given by an optional parameter name. It can be used to identify this pyEval operator in debug output, or in the dryrun mode of simulator::evolve.

2.11.5 Class pyExec (Function form: PyExec)

Execute a Python statement This operator takes a list of statements and executes them. No value will be returned or outputted.

class pyExec(*stmts="", preStmts="", postStmts="", exposePop=False, name="", output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
evaluate statments in the local replicate namespace, no return value Please refer to class pyEval for parameter descriptions.

clone()
Deep copy of a pyExec operator

2.11.6 Class infoEval (Function form: infoEval)

Unlike operator pyEval and pyExec that work at the population level, in its local namespace, infoEval works at the individual level, working with individual information fields. is statement can change the value of existing information fields. Optionally, variables in population's local namespace can be used in the statement, but this should be used with caution.

class infoEval(*expr="", stmts="", subPops=[], usePopVars=False, exposePop=False, name="", output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

exposePop: If `True`, expose the current population as a variable named `pop`

expr: The expression to be evaluated. The result will be sent to `output`.

name: Used to let pure Python operator to identify themselves

output: Default to >. I.e., output to standard output. Note that because the expression will be executed for each individual, the output can be large.

stmts: The statement that will be executed before the expression

subPop: A shortcut to `subPops=[subPop]`

subPops: Subpopulations this operator will apply to. Default to all.

usePopVars: If `True`, import variables from expose the current population as a variable named `pop`

evaluate Python statements with variables being an individual's information fields The expression and statements will be executed for each individual, in a Python namespace (dictionary) where individual information fields are made available as variables. Population dictionary can be made avaialbe with option usePopVars. Changes to these variables will change the corresponding information fields of individuals.Please note that, 1. If population variables are used, and there are name conflicts between information fields and variables, population variables will be overridden by information fields, without any warning. 2. Information fields are float numbers. An exceptions will raise if an information field can not be converted to a float number. 3. This operator can be used in all stages. When it is used during-mating, it will act on each offspring.

apply(*pop*)
Apply the `infoEval` operator

clone()
Deep copy of a `infoEval` operator

name()
Return the name of an expression The name of a `infoEval` operator is given by an optional parameter `name`. It can be used to identify this `infoEval` operator in debug output, or in the dryrun mode of simulator::evolve.

2.11.7 Class `infoExec` (Function form: `infoExec`)

Execute a Python statement for each individual, using information fields This operator takes a list of statements and executes them. No value will be returned or outputted.

class infoExec(*stmts="", subPops=[], usePopVars=False, exposePop=False, name="", output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
fields, optionally with variable in population's local namespace Please refer to class `infoEval` for parameter descriptions.

clone()
Deep copy of a `infoExec` operator

2.12 Tagging (used for pedigree tracking)

2.12.1 Class `tagger`

Base class of tagging individuals This is a during-mating operator that tags individuals with various information. Potential usages are:

- recording the parental information to track pedigree;

- tagging an individual/allele and monitoring its spread in the population etc.

class tagger(*output="", outputExpr="", begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
create a `tagger`, default to be always active but no output

apply(*pop*)
Add a newline

clone()
Deep copy of a
tagger

2.12.2 Class `inheritTagger`

Inherite tag from parents This during-mating operator will copy the tag (information field) from his/her parents. Depending on `mode` parameter, this tagger will obtain tag, value of the first specified information fields, from his/her father or mother (two tag fields), or both (first tag field from father, and second tag field from mother).

An example may be tagging one or a few parents and examining, at the last generation, how many offspring they have.

class inheritTagger(*mode=TAG_Paternal, begin=0, end=-1, step=1, at=[], rep=REP_ALL, output="", outputExpr="", infoFields=["paternal_tag", "maternal_tag"]*)
mode: Can be one of `TAG_Paternal`, `TAG_Maternal`, and `TAG_Both`

create an `inheritTagger` that inherits a tag from one or both parents

applyDuringMating(*pop, offspring, dad=None, mom=None*)
Apply the `inheritTagger`

clone()
Deep copy of a `inheritTagger`

2.12.3 Class `parentTagger`

Tagging according to parental indexes This during-mating operator set `tag()` each individual with indexes of his/her parent in the parental population. Because only one parent is recorded, this is recommended to be used for mating schemes that requires only one parent (such as selfMating).This tagger record indexes to information field parent_idx, and/or a given file. The usage is similar to parentsTagger.

class parentTagger(*begin=0, end=-1, step=1, at=[], rep=REP_ALL, output="", outputExpr="", infoFields=["parent_idx"]*)
create a `parentTagger`

apply(*pop*)
With a newline.

applyDuringMating(*pop, offspring, dad=None, mom=None*)
Apply the `parentTagger`

clone()
Deep copy of a `parentTagger`

2.12.4 Class `parentsTagger`

Tagging according to parents' indexes This during-mating operator set `tag()`, currently a pair of numbers, of each individual with indexes of his/her parents in the parental population. This information will be used by pedigree-related

operators like `affectedSibpairSample` to track the pedigree information. Because parental population will be discarded or stored after mating, these index will not be affected by post-mating operators.This tagger record parental index to one or both

- one or two information fields. Default to father_idx and mother_idx. If only one parent is passed in a mating scheme (such as selfing), only the first information field is used. If two parents are passed, the first information field records paternal index, and the second records maternal index.
- a file. Indexes will be written to this file. This tagger will also act as a post-mating operator to add a new-line to this file.

class parentsTagger(*begin=0, end=-1, step=1, at=[], rep=REP_ALL, output="", outputExpr="", infoFields=["father_idx", "mother_idx"]*)

create a `parentsTagger`

apply(*pop*)

With a newline.

applyDuringMating(*pop, offspring, dad=None, mom=None*)

Apply the `parentsTagger`

clone()

Deep copy of a `parentsTagger`

2.12.5 Class `sexTagger`

Tagging sex status. This is a simple post-mating tagger that write sex status to a file. By default, 1 for Male, 2 for Female.

class sexTagger(*code=[], begin=0, end=-1, step=1, at=[], rep=REP_ALL, stage=PostMating, output=">", outputExpr="", infoFields=[]*)

code: Code for Male and Female, default to 1 and 2, respectively. This is used by Linkage format.

FIXME: No document

apply(*pop*)

Add a newline

2.12.6 Class `affectionTagger`

Tagging affection status. This is a simple post-mating tagger that write affection status to a file. By default, 1 for unaffected, 2 for affected.

class affectionTagger(*code=[], begin=0, end=-1, step=1, at=[], rep=REP_ALL, stage=PostMating, output=">", outputExpr="", infoFields=[]*)

code: Code for Male and Female, default to 1 and 2, respectively. This is used by Linkage format.

FIXME: No document

apply(*pop*)

Add a newline

2.12.7 Class `infoTagger`

Tagging information fields. This is a simple post-mating tagger that write given information fields to a file (or standard output).

class infoTagger(*begin=0, end=-1, step=1, at=[], rep=REP_ALL, stage=PostMating, output=">", outputExpr="", infoFields=[]*)

FIXME: No document

apply(*pop*)

Add a newline

2.12.8 Class pyTagger

Python tagger. This tagger takes some information fields from both parents, pass to a Python function and set the individual field with the return value. This operator can be used to trace the inheritance of trait values.

class pyTagger(*func=None, begin=0, end=-1, step=1, at=[], rep=REP_ALL, output="", outputExpr="", infoFields=[]*)

func: A Pyton function that returns a list to assign the information fields. e.g., if `fields=['A', 'B']`, the function will pass values of fields `'A'` and `'B'` of father, followed by mother if there is one, to this function. The return value is assigned to fields `'A'` and `'B'` of the offspring. The return value has to be a list even if only one field is given.

infoFields: Information fields. The user should gurantee the existence of these fields.

creates a `pyTagger` that works on specified information fields

applyDuringMating(*pop, offspring, dad=None, mom=None*)

Apply the `pyTagger`

clone()

Deep copy of a `pyTagger`

2.13 Terminator

2.13.1 Class terminator

Base class of all terminators. Teminators are used to see if an evolution is running as expected, and terminate the evolution if a certain condition fails.

class terminator(*message="", output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

message: A message that will be displayed when the evolution is terminated.

create a terminator

clone()

Deep copy of a terminator

2.13.2 Class terminateIf

Terminate according to a condition This operator terminates the evolution under certain conditions. For example, `terminateIf(condition='alleleFreq[0][1]<0.05', begin=100)` terminates the evolution if the allele frequency of allele `1` at locus `0` is less than 0.05. Of course, to make this opertor work, you will need to use a `stat` operator before it so that variable `alleleFreq` exists in the local namespace.

When the value of condition is `True`, a shared variable `var="terminate"` will be set to the current generation.

class terminateIf(*condition="", message="", var="terminate", output="", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

create a `terminateIf` terminator

apply(*pop*)
Apply the `terminateIf` terminator

clone()
Deep copy of a `terminateIf` terminator

2.13.3 Class `continueIf`

Terminate according to a condition failure The same as `terminateIf` but continue if the condition is `True`.

class continueIf(*condition="", message="", var="terminate", output="", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
create a `continueIf` terminator

apply(*pop*)
Apply this operator

clone()
Deep copy of a `continueIf` terminator

2.14 Python operators

2.14.1 Class `pyOperator`

A python operator that directly operate a population. This operator accepts a function that can take the form of

- `func(pop)` when `stage=PreMating` or `PostMating`, without setting `param`;
- `func(pop, param)` when `stage=PreMating` or `PostMating`, with `param`;
- `func(pop, off, dad, mom)` when `stage=DuringMating` and `passOffspringOnly=False`, without setting `param`;
- `func(off)` when `stage=DuringMating` and `passOffspringOnly=True`, and without setting `param`;
- `func(pop, off, dad, mom, param)` when `stage=DuringMating` and `passOffspringOnly=False`, with `param`;
- `func(off, param)` when `stage=DuringMating` and `passOffspringOnly=True`, with `param`.

For `Pre-` and `PostMating` usages, a population and an optional parameter is passed to the given function. For `DuringMating` usages, population, offspring, its parents and an optional parameter are passed to the given function. Arbitrary operations can be applied to the population and offspring (if `stage=DuringMating`).

class pyOperator(*func, param=None, stage=PostMating, formOffGenotype=False, passOffspringOnly=False, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
formOffGenotype: This option tells the mating scheme this operator will set the genotype of offspring (valid only for `stage=DuringMating`). By default (`formOffGenotype=False`), a mating scheme will set the genotype of offspring before it is passed to the given Python function. Otherwise, a 'blank' offspring will be passed.

func: A Python function. Its form is determined by other parameters.

param: Any Python object that will be passed to `func` after `pop` parameter. Multiple parameters can be passed as a tuple.

passOffspringOnly: If `True`, `pyOperator` will expect a function of form `func(off [,param])`, instead of `func(pop, off, dad, mom [, param])` which is used when `passOffspringOnly` is `False`. Because many during-mating `pyOperator` only need access to offspring, this will improve efficiency. Default to `False`.

Python operator, using a function that accepts a population object.

Note

- Output to `output` or `outputExpr` is not supported. That is to say, you have to open/close/append to files explicitly in the Python function. Because files specified by `output` or `outputExpr` are controlled (opened/closed) by simulators, they should not be manipulated in a `pyOperator` operator.
- This operator can be applied `Pre-`, `During-` or `Post- Mating` and is applied `PostMating` by default. For example, if you would like to examine the fitness values set by a selector, a `PreMating` Python operator should be used.

`apply`(*pop*)
Apply the `pyOperator` operator to one population

`clone`()
Deep copy of a `pyOperator` operator

2.14.2 Class `pyIndOperator`

Individual operator This operator is similar to a `pyOperator` but works at the individual level. It expects a function that accepts an individual, optional genotype at certain loci, and an optional parameter. When it is applied, it passes each individual to this function. When `infoFields` is given, this function should return an array to fill these `infoFields`. Otherwise, `True` or `False` is expected.More specifically, `func` can be

- `func(ind)` when neither `loci` nor `param` is given.
- `func(ind, genotype)` when `loci` is given.
- `func(ind, param)` when `param` is given.
- `func(ind, genotype, param)` when both `loci` and `param` are given.

class `pyIndOperator`(*func, loci=[], param=None, stage=PostMating, formOffGenotype=False, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
func: A Python function that accepts an individual and optional genotype and parameters.

infoFields: If given, `func` is expected to return an array of the same length and fill these `infoFields` of an individual.

param: Any Python object that will be passed to `func` after `pop` parameter. Multiple parameters can be passed as a tuple.

a `Pre-` or `PostMating` Python operator that apply a function to each individual

`apply`(*pop*)
Apply the `pyIndOperator` operator to one population

`clone`()
Deep copy of a `pyIndOperator` operator

2.15 Miscellaneous

2.15.1 Class `ifElse`

Conditional operator This operator accepts

- an expression that will be evaluated when this operator is applied.
- an operator that will be applied if the expression is `True` (default to null).
- an operator that will be applied if the expression is `False` (default to null).

When this operator is applied to a population, it will evaluate the expression and depending on its value, apply the supplied operator. Note that the `begin`, `end`, `step`, and `at` parameters of `ifOp` and `elseOp` will be ignored. For example, you can mimic the `at` parameter of an operator by `ifElse('rep in [2,5,9]' operator)`. The real use of this machanism is to monitor the population statistics and act accordingly.

class `ifElse`(*cond, ifOp=None, elseOp=None, output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

cond: Expression that will be treated as a boolean variable

elseOp: An operator that will be applied when `cond` is `False`

ifOp: An operator that will be applied when `cond` is `True`

create a conditional operator

`apply`(*pop*)

Apply the `ifElse` operator to one population

`clone`()

Deep copy of an `ifElse` operator

2.15.2 Class `turnOnDebug` (Function form: `TurnOnDebug`)

Set debug on Turn on debug. There are several ways to turn on debug information for non-optimized modules, namely

- set environment variable `SIMUDEBUG`.
- use `simuOpt.setOptions(debug)` function.
- use `TurnOnDebug` or `TurnOnDebugByName` function.
- use this `turnOnDebug` operator

The advantage of using this operator is that you can turn on debug at given generations.

class `turnOnDebug`(*code, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

create a `turnOnDebug` operator

`apply`(*pop*)

Apply the `turnOnDebug` operator to one population

`clone`()

Deep copy of a `turnOnDebug` operator

2.15.3 Class `turnOffDebug` (Function form: `TurnOffDebug`)

Set debug off Turn off debug.

class `turnOffDebug`(*code, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)

create a `turnOffDebug` operator

`apply`(*pop*)

Apply the `turnOffDebug` operator to one population

`clone`()

Deep copy of a `turnOffDebug` operator

2.15.4 Class noneOp

None operator This operator does nothing.

class noneOp(*output=">", outputExpr="", stage=PostMating, begin=0, end=0, step=1, at=[], rep=REP_ALL, infoFields=[]*)
create a none operator

apply(*pop*)
Apply the noneOp operator to one population

clone()
Deep copy of a noneOp operator

2.15.5 Class pause

Pause a simulator This operator pauses the evolution of a simulator at given generations or at a key stroke, using stopOnKeyStroke=True option. Users can use 'q' to stop an evolution. When a simulator is stopped, press any other key to resume the simulation or escape to a Python shell to examine the status of the simulation by pressing 's'.

There are two ways to use this operator, the first one is to pause the simulation at specified generations, using the usual operator parameters such as at. Another way is to pause a simulation with any key stroke, using the stopOnKeyStroke parameter. This feature is useful for a presentation or an interactive simulation. When 's' is pressed, this operator expose the current population to the main Python dictionary as variable pop and enter an interactive Python session. The way current population is exposed can be controlled by parameter exposePop and popName. This feature is useful when you want to examine the properties of a population during evolution.

class pause(*prompt=True, stopOnKeyStroke=False, exposePop=True, popName="pop", output=">", outputExpr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=REP_LAST, infoFields=[]*)
exposePop: Whether or not expose pop to user namespace, only useful when user choose 's' at pause. Default to True.

popName: By which name the population is exposed. Default to pop.

prompt: If True (default), print prompt message.

stopOnKeyStroke: If True, stop only when a key was pressed.

stop a simulation. Press 'q' to exit or any other key to continue.

apply(*pop*)
Apply the pause operator to one population

clone()
Deep copy of a pause operator

2.15.6 Class ticToc (Function form: TicToc)

Timer operator This operator, when called, output the difference between current and the last called clock time. This can be used to estimate execution time of each generation. Similar information can also be obtained from turnOnDebug(DBG_PROFILE), but this operator has the advantage of measuring the duration between several generations by setting step parameter.

class ticToc(*output=">", outputExpr="", stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
create a timer

apply(*pop*)
Apply the ticToc operator to one population

clone()
Deep copy of a `ticToc` operator

2.15.7 Class `setAncestralDepth`

Set ancestral depth This operator set the number of ancestral generations to keep in a population. It is usually called like `setAncestral(at=[-2])` to start recording ancestral generations to a population at the end of the evolution. This is useful when constructing pedigree trees from a population.

class setAncestralDepth(*depth, output=">", outputExpr="", stage=PreMating, begin=0, end=-1, step=1, at=[], rep=REP_ALL, infoFields=[]*)
create a `setAncestralDepth` operator

apply(*pop*)
Apply the `setAncestralDepth` operator to one population

clone()
Deep copy of a `setAncestralDepth` operator

CHAPTER

THREE

Global and Python Utility functions

3.1 Global functions

AlleleType()
Return the allele type of the current module. Can be `binary`, `short`, or `long`.

Limits()
Print out system limits

ListAllRNG()
List the names of all available random number generators

ListDebugCode()
List all debug codes

LoadPopulation(*file, format="auto"*)
Load a population from a file. The file format is by default determined by file extension (`format="auto"`). Otherwise, `format` can be one of `txt`, `bin`, or `xml`.

LoadSimulator(*file, mate, format="auto"*)
Load a simulator from a file with the specified mating scheme. The file format is by default determined by file extension (`format="auto"`). Otherwise, `format` can be one of `txt`, `bin`, or `xml`.

MaxAllele()
Return 1, $2^8 - 1$, $2^{16} - 1$ for binary, short, or long allele modules, respectively

MergePopulations(*pops, newSubPopSizes=[], keepAncestralPops=-1*)
Merge several populations with the same genotypic structure and create a new population

MergePopulationsByLoci(*pops, newNumLoci=[], newLociPos=[], byChromosome=False*)
Merge several populations of the same size by loci and create a new population

ModuleCompiler()
Return the compiler used to compile this simuPOP module

ModuleDate()
Return the date when this simuPOP module is compiled

ModulePlatForm()
Return the platform on which this simuPOP module is compiled

ModulePyVersion()
Return the Python version this simuPOP module is compiled for

Optimized()
Return `True` if this simuPOP module is optimized

SetRNG(*rng="", seed=0*)

Set random number generator. If `seed=0` (default), a random seed will be given. If `rng=""`, seed will be set to the current random number generator.

TurnOffDebug(*code=DBG_ALL*)
Turn off debug information. Default to turn off all debug codes. Only available in non-optimized modules.

TurnOnDebug(*code=DBG_ALL*)
Set debug codes. Default to turn on all debug codes. Only available in non-optimized modules.

rng()
Return the currently used random number generator

simuRev()
Return the revision number of this simuPOP module. Can be used to test if a feature is available.

simuVer()
Return the version of this simuPOP module

3.2 Utility Classes

3.2.1 Class `RNG`

Random number generator This random number generator class wraps around a number of random number generators from GNU Scientific Library. You can obtain and change system random number generator through the `rng()` function. Or create a separate random number generator and use it in your script.

class RNG(*rng=None, seed=0*)
RNG used by simuPOP.

max()
Maximum value of this RNG.

maxSeed()
Return the maximum allowed seed value

name()
Return RNG name

pvalChiSq(*chisq, df*)
Right hand side (single side) p-value for ChiSq value

randBinomial(*n, p*)
Binomial distribution B(n, p).

randExponential(*v*)
FIXME: No document

randGeometric(*p*)
Geometric distribution.

randGet()
Return a random number in the range of [0, 2, ... max()-1]

randInt(*n*)
Return a random number in the range of [0, 1, 2, ... n-1]

randMultinomial(*N, p, n*)
Multinomial distribution.

randMultinomialVal(*N, p*)
FIXME: No document

randNormal(*m, v*)
Normal distribution.

randPoisson(*p*)
Poisson distribution.

randUniform01()
Uniform distribution [0,1).

seed()
Return the seed of this RNG

setRNG(*rng=None, seed=0*)
Choose an random number generator, or set seed to the current RNG

rng: Name of the RNG. If rng is not given, environmental variable GSL_RNG_TYPE will be used if it is available. Otherwise, RNG`mt19937` will be used.

seed: Random seed. If not given, `/dev/urandom`, `/dev/random`, system time will be used, depending on availability, in that order. Note that windows system does not have `/dev` so system time is used.

setSeed(*seed*)
If seed is 0, method described in `setRNG` is used.

3.3 Utility Modules

Several utility modules are distributed with simuPOP. They provide important functions and extensions to simuPOP and serve as good examples on how simuPOP can be used.

Compared to simuPOP kernel functions, these utility functions are less tested, and are subject to more frequent changes. Please report to simuPOP mailing list if any function stops working.

3.3.1 Module `simuOpt`

Module simuOpt can be used to control which simuPOP module to load, and how it is loaded using function setOptions . It also provides a simple way to set simulation options, from user input, command line, configuration file or a parameter dialog. All you need to do is to define an option description list that lists all parameters in a given format, and call the getParam function.

This module, if loaded, pre-process the command line options. More specifically, it checks command line option:

-c configfile: read from a configuration file

--config configfile: the same as -c

--optimized: load optimized modules, unless setOption explicitly use non-optimized: modules.

-q: Do not display banner information when simuPOP is loaded

--quiet: the same as -q

--useTkinter: force the use of Tcl/Tk dialog even when wxPython is available. By: default, wxPython is used whenever possible.

--noDialog: do not use option dialog. If the options can not be obtained from: command line or configuraiton file, users will be asked to input them interactively.

Because these options are reserved, you can not use them in your simuPOP script.

Module Functions

getParam(*options=[], doc='', details='', noDialog=False, UnprocessedArgs=True, verbose=False, nCol=1*)
Get parameters from either:

- a Tcl/Tk based, or wxPython based parameter dialog (wxPython is used if it is available)
- command line argument
- configuration file specified by -c file (--config file), or
- prompt for user input

The option description list consists of dictionaries with some predefined keys. Each dictionary defines an option. Each option description item can have the following keys:

arg: short command line option name. 'h' checks the presence of argument -h . If an argument is expected, add a comma to the option name. For example, 'p:' matches command line option -p=100 or -p 100 .

longarg: long command line option name. 'help' checks the presence of: argument '--help' . 'mu=' matches command line option --mu=0.001 or -mu 0.001 .

label: The label of the input field in a parameter dialog, and as the prompt for: user input.

default: default value for this parameter. It is used to as the default value: in the parameter dialog, and as the option value when a user presses 'Enter' directly during interactive parameter input.

useDefault: use default value without asking, if the value can not be determined: from GUI, command line option or config file. This is useful for options that rarely need to be changed. Setting them to useDfault allows shorter command lines, and easy user input.

description: a long description of this parameter, will be put into the usage: information, which will be displayed with (-h , --help command line option, or help button in parameter dialog).

allowedTypes: acceptable types of this option. If allowedTypes is types.ListType: or types.TupleType and the user's input is a scalar, the input will be converted to a list automatically. If the conversion can not be done, this option will not be accepted.

validate: a function to validate the parameter. You can define your own functions: or use the ones defined in this module.

chooseOneOf: if specified, simuOpt will choose one from a list of values using a: listbox (Tk) or a combo box (wxPython) .

chooseFrom: if specified, simuOpt will choose one or more items from a list of: values using a listbox (tk) or a combo box (wxPython).

separator: if specified, a blue label will be used to separate groups of: parameters.

jump: it is used to skip some parameters when doing the interactive user input.: For example, getParam will skip the rest of the parameters if -h is specified if parameter -h has item 'jump':-1 which means jumping to the end. Another situation of using this value is when you have a hierarchical parameter set. For example, if mutation is on, specify mutation rate, otherwise proceed. The value of this option can be the absolute index or the longarg name of another option.

jumpIfFalse: The same as jump but jump if current parameter is False .

This function will first check command line argument. If the argument is available, use its value. Otherwise check if a config file is specified. If so, get the value from the config file. If both failed, prompt user to input a value. All input will be checked against types, if exists, an array of allowed types.

Parameters of this function are:

options: a list of option description dictionaries

doc: short description put to the top of parameter dialog

details: module help. Usually set to __doc__ .

noDialog: do not use a parameter dialog, used in batch mode. Default to False.

checkUnprocessedArgs: obsolete because unused args are always checked.

verbose: whether or not print detailed info

nCol: number of columns in the parameter dialog.

prettyOutput(*value, quoted=False, outer=True*)
Return a value in good format, the main purpose is to avoid [0.90000001, 0.2].

printConfig(*opt, param, out=<open file '<stdout>', mode 'w' at 0x2aaaaaad6198>*)
Print configuration.

opt: option description list

param: parameters returned from getParam()

out: output

requireRevision(*rev*)
Compare the revision of this simuPOP module with given revision. Raise an exception if current module is out of date.

saveConfig(*opt, file, param*)
Write a configuration file. This file can be later read with command line option -c or --config .

opt: the option description list

file: output file

param: parameters returned from getParam

setOptions(*optimized=None, mpi=None, chromMap=[], alleleType=None, quiet=None, debug=[]*)
set options before simuPOP is loaded to control which simuPOP module to load, and how the module should be loaded.

optimized: whether or not load optimized version of a module. If not set,: environmental variable SIMUOPTIMIZED, and commandline option --optimized will be used if available. If nothing is defined, standard version will be used.

mpi: obsolete.

chromMap: obsolete.

alleleType: 'binary', 'short', or 'long'. 'standard' can be used as 'short': for backward compatibility. If not set, environmental variable SIMUALLELETYPE will be used if available. if it is not defined, the short allele version will be used.

quiet: If True, supress banner information when simuPOP is loaded.

debug: a list of debug code (or string). If not set, environmental variable: SIMUDEBUG will be used if available.

usage(*options, before=''*)
Print usage information from the option description list. Used with -h (or --help) option, and in the parameter input dialog.

options: option description list.

before: optional information

valueAnd(*t1, t2*)
Return a function that returns true if passed option passes validator t1 and t2

valueBetween(*a, b*)
Return a function that returns true if passed option is between value a and b (a and b included)

valueEqual(*a*)
Return a function that returns true if passed option equals a

valueGE(*a*)
Return a function that returns true if passed option is greater than or equal to a

valueGT(*a*)
Return a function that returns true if passed option is greater than a

valueIsList()
Return a function that returns true if passed option is a list (or tuple)

valueIsNum()
Return a function that returns true if passed option is a number (int, long or float)

valueLE(*a*)
Return a function that returns true if passed option is less than or equal to a

valueLT(*a*)
Return a function that returns true if passed option is less than a

valueListOf(*t*)
Return a function that returns true if passed option val is a list of type t. If t is a function (validator), check if all v in val pass t(v)

valueNot(*t*)
Return a function that returns true if passed option does not passes validator t

valueNotEqual(*a*)
Return a function that returns true if passed option does not equal a

valueOneOf(*t*)
Return a function that returns true if passed option is one of the values list in t

valueOr(*t1, t2*)
Return a function that returns true if passed option passes validator t1 or t2

valueTrueFalse()
Return a function that returns true if passed option is True or False

valueValidDir()
Return a function that returns true if passed option val if a valid directory

valueValidFile()
Return a function that returns true if passed option val if a valid file

3.3.2 Module `simuUtil`

This module provides some commonly used operators and format conversion utilities.

Module Functions

CaseControl_ChiSq(*pop, sampleSize, penetrance=None*)
Draw affected sibpair sample from pop, run TDT using GENEHUNTER

pop: simuPOP population. It can be a string if path to a file is given. This population must

- have at least one ancestral generation (parental generation)
- have a variable DSL (pop.dvars().DSL) indicating the Disease susceptibility loci. These DSL will be removed from the samples.
- has only binary alleles

pene: penetrance function, if not given (None), existing affection status will be used.

sampleSize: total sample size N. N/4 is the number of families to ascertain.

keep_temp: if True, do not remove sample data. Default to False.

ChiSq_test(*pop*)
perform case control test

pop: loaded population, or population file in simuPOP format. This function assumes that pop has two subpopulations, cases and controls, and have 0 as wildtype and 1 as disease allele. pop can also be an loaded

population object.

Return value: A list of p-value at each locus.

Note: this function requires rpy module.

ConstSize (*size, split=0, numSubPop=1, bottleneckGen=-1, bottleneckSize=0*)
The population size is constant, but will split into numSubPop subpopulations at generation split

ExponentialExpansion (*initSize, endSize, end, burnin=0, split=0, numSubPop=1, bottleneckGen=-1, bottleneckSize=0*)
Exponentially expand population size from intiSize to endSize after burnin, split the population at generation split.

InstantExpansion (*initSize, endSize, end, burnin=0, split=0, numSubPop=1, bottleneckGen=-1, bottleneckSize=0*)
Instaneously expand population size from intiSize to endSize after burnin, split the population at generation split.

LOD_gh (*file, gh='gh'*)
Analyze data using the linkage method of genehunter. Note that this function may not work under platforms other than linux, and may not work with your version of genehunter. As a matter of fact, it is almost unrelated to simuPOP and is provided only as an example how to use python to analyze data.

Parameters: file: file to analyze. This function will look for file.dat and file.pre in linkage format.

loci: a list of loci at which p-value will be returned. If the list is empty, all p-values are returned.

gh: name (or full path) of genehunter executable. Default to 'gh'

Return value: A list (for each chromosome) of list (for each locus) of p-values.

LOD_merlin (*file, merlin='merlin'*)
run multi-point non-parametric linkage analysis using merlin

LargePeds_Reg_merlin (*pop, sampleSize, qtrait=None, infoField='qtrait', merlin='merlin-regress', keep_temp=False*)
Draw affected sibpair sample from pop, run TDT using GENEHUNTER

pop: simuPOP population. It can be a string if path to a file is given. This population must

- have at least one ancestral generation (parental generation)
- have a variable DSL (pop.dvars().DSL) indicating the Disease susceptibility loci. These DSL will be removed from the samples.
- has only binary alleles

qtrait: a function to calculate quantitative trait

infoField: information field to store quantitative trait. Default to 'qtrait'

sampleSize: total sample size N. N/4 is the number of families to ascertain.

merlin: executable name of merlin, full path name can be given.

keep_temp: if True, do not remove sample data. Default to False.

LargePeds_VC_merlin (*pop, sampleSize, qtrait=None, infoField='qtrait', merlin='merlin', keep_temp=False*)
Draw affected sibpair sample from pop, run TDT using GENEHUNTER

pop: simuPOP population. It can be a string if path to a file is given. This population must

- have at least one ancestral generation (parental generation)
- have a variable DSL (pop.dvars().DSL) indicating the Disease susceptibility loci. These DSL will be removed from the samples.
- has only binary alleles

qtrait: a function to calculate quantitative trait

infoField: information field to store quantitative trait. Default to 'qtrait'

sampleSize: total sample size N. N/4 is the number of families to ascertain.

merlin: executable name of merlin, full path name can be given.

keep_temp: if True, do not remove sample data. Default to False.

LinearExpansion(*initSize, endSize, end, burnin=0, split=0, numSubPop=1, bottleneckGen=-1, bottleneckSize=0*)

Linearly expand population size from intiSize to endSize after burnin, split the population at generation split.

ListVars(*var, level=-1, name='', subPop=True, useWxPython=True*)

list a variable in tree format, either in text format or in a: wxPython window.

var: any variable to be viewed. Can be a dw object returned by dvars() function

level: level of display.

name: only view certain variable

subPop: whether or not display info in subPop

useWxPython: if True, use terminal output even if wxPython is available.

LoadFstat(*file, loci=[]*)

load population from fstat file 'file' since fstat does not have chromosome structure an additional parameter can be given

MigrIslandRates(*r, n*)

migration rate matrix

```
x m/(n-1) m/(n-1) ....
m/(n-1) x .............
.....
.... m/(n-1) m/(n-1) x
```

where x = 1-m

MigrSteppingStoneRates(*r, n, circular=False*)

migration rate matrix, circular stepping stone model (X=1-m)

```
X   m/2               m/2
m/2 X   m/2           0
0   m/2 x   m/2 ......0
...
m/2 0 ....        m/2  X
```

or non-circular

```
X   m/2               m/2
m/2 X   m/2           0
0   m/2 X   m/2 ......0
...
...               m   X
```

QtraitSibs_Reg_merlin(*pop, sampleSize, qtrait=None, infoField='qtrait', merlin='merlin-regress', keep_temp=False*)

Draw affected sibpair sample from pop, run TDT using GENEHUNTER

pop: simuPOP population. It can be a string if path to a file is given. This population must

- have at least one ancestral generation (parental generation)

- have a variable DSL (pop.dvars().DSL) indicating the Disease susceptibility loci. These DSL will be removed from the samples.
- has only binary alleles

qtrait: a function to calculate quantitative trait

infoField: information field to store quantitative trait. Default to 'qtrait'

sampleSize: total sample size N. N/4 is the number of families to ascertain.

merlin: executable name of merlin, full path name can be given.

keep_temp: if True, do not remove sample data. Default to False.

QtraitSibs_VC_merlin (*pop, sampleSize, qtrait=None, infoField='qtrait', merlin='merlin', keep_temp=False*)

Draw affected sibpair sample from pop, run TDT using GENEHUNTER

pop: simuPOP population. It can be a string if path to a file is given. This population must

- have at least one ancestral generation (parental generation)
- have a variable DSL (pop.dvars().DSL) indicating the Disease susceptibility loci. These DSL will be removed from the samples.
- has only binary alleles

qtrait: a function to calculate quantitative trait

infoField: information field to store quantitative trait. Default to 'qtrait'

sampleSize: total sample size N. N/4 is the number of families to ascertain.

merlin: executable name of merlin, full path name can be given.

keep_temp: if True, do not remove sample data. Default to False.

Regression_merlin (*file, merlin='merlin-regress'*)

run merlin regression method

SaveCSV (*pop, output='', outputExpr='', fields=['sex', 'affection'], loci=[], combine=None, shift=1, **kwargs*)

save file in CSV format

fileds: information fields, 'sex' and 'affection' are special fields that is treated differently.

genotype: list of loci to output, default to all.

combine: how to combine the markers. Default to None. A function can be specified, that takes the form:

```
def func(markers):
    return markers[0]+markers[1]
```

shift: since alleles in simuPOP is 0-based, shift=1 is usually needed to output alleles starting from allele 1. This parameter is ignored if combine is used.

SaveFstat (*pop, output='', outputExpr='', maxAllele=0, loci=[], shift=1, combine=None*)

save file in FSTAT format

SaveLinkage (*pop, output='', outputExpr='', loci=[], shift=1, combine=None, fields=[], recombination=1.0000000000000001e-05, penetrance=[0, 0.25, 0.5], affectionCode=['1', '2'], pre=True, daf=0.001*)

save population in Linkage format. Currently only support affected sibpairs sampled with affectedSibpairSample operator.

pop: population to be saved. Must have ancestralDepth 1. paired individuals are sibs. Parental population are corresponding parents. If pop is a filename, it will be loaded.

output: Output.dat and output.ped will be the data and pedigree file. You may need to rename them to be analyzed by LINKAGE. This allows saving multiple files.

outputExpr: expression version of output.

affectionCode: default to '1': unaffected, '2': affected

pre: True. pedigree format to be fed to makeped. Non-pre format it is likely to be wrong now for non-sibpair families.

Note: the first child is always the proband.

SaveMerlinDatFile(*pop, output='', outputExpr='', loci=[], fields=[], outputAffection=False*)
Output a .dat file readable by merlin

SaveMerlinMapFile(*pop, output='', outputExpr='', loci=[]*)
Output a .map file readable by merlin

SaveMerlinPedFile(*pop, output='', outputExpr='', loci=[], fields=[], header=False, outputAffection=False, affectionCode=['U', 'A'], combine=None, shift=1, **kwargs*)
Output a .ped file readable by merlin

SaveQTDT(*pop, output='', outputExpr='', loci=[], header=False, affectionCode=['U', 'A'], fields=[], combine=None, shift=1, **kwargs*)
save population in Merlin/QTDT format. The population must have pedindex, father_idx and mother_idx information fields.

pop: population to be saved. If pop is a filename, it will be loaded.

output: base filename.

outputExpr: expression for base filename, will be evaluated in pop's local namespace.

affectionCode: code for unaffected and affected. '1', '2' are default, but 'U', and 'A' or others can be specified.

loci: loci to output

header: whether or not put head line in the ped file.

fields: information fields to output

combine: an optional function to combine two alleles of a diploid individual.

shift: if combine is not given, output two alleles directly, adding this value (default to 1).

SaveSolarFrqFile(*pop, output='', outputExpr='', loci=[], calcFreq=True*)
Output a frequency file, in a format readable by solar calcFreq

Unexpected indentation.

whether or not calculate allele frequency

Sibpair_LOD_gh(*pop, sampleSize, penetrance=None, recRate=None, daf=None, gh='gh', keep_temp=False*)
Draw affected sibpair sample from pop, run Linkage analysis using GENEHUNTER

pop: simuPOP population. It can be a string if path to a file is given. This population must

- have at least one ancestral generation (parental generation)
- have a variable DSL (pop.dvars().DSL) indicating the Disease susceptibility loci. These DSL will be removed from the samples.
- has only binary alleles

pene: penetrance function, if not given (None), existing affection status will be used.

sampleSize: total sample size N. N/4 is the number of families to ascertain.

recRate: recombination rate, used in the Linkage file. If not given, pop.dvars().recRate[0] will be used. If there is no such variable, 0.0001 is used.

daf: disease allele frequency. This is needed for the linkage format but I am not sure if it is used by TDT.

gh: executable name of genehunter, full path name can be given.

keep_temp: if True, do not remove sample data. Default to False.

Sibpair_LOD_merlin (*pop, sampleSize, penetrance=None, merlin='merlin', keep_temp=False*)
Draw affected sibpair sample from pop, run multi-point linkage analysis using merlin

pop: simuPOP population. It can be a string if path to a file is given. This population must

- have at least one ancestral generation (parental generation)
- have a variable DSL (pop.dvars().DSL) indicating the Disease susceptibility loci. These DSL will be removed from the samples.
- has only binary alleles

pene: penetrance function, if not given (None), existing affection status will be used.

sampleSize: total sample size N. N/4 is the number of families to ascertain.

merlin: executable name of merlin, full path name can be given.

keep_temp: if True, do not remove sample data. Default to False.

Sibpair_TDT_gh (*pop, sampleSize, penetrance=None, recRate=None, daf=None, gh='gh', keep_temp=False*)
Draw affected sibpair sample from pop, run TDT using GENEHUNTER

pop: simuPOP population. It can be a string if path to a file is given. This population must

- have at least one ancestral generation (parental generation)
- have a variable DSL (pop.dvars().DSL) indicating the Disease susceptibility loci. These DSL will be removed from the samples.
- has only binary alleles

pene: penetrance function, if not given (None), existing affection status will be used.

sampleSize: total sample size N. N/4 is the number of families to ascertain.

recRate: recombination rate, used in the Linkage file. If not given, pop.dvars().recRate[0] will be used. If there is no such variable, 0.0001 is used.

daf: disease allele frequency. This is needed for the linkage format but I am not sure if it is used by TDT.

gh: executable name of genehunter, full path name can be given.

keep_temp: if True, do not remove sample data. Default to False.

TDT_gh (*file, gh='gh'*)
Analyze data using genehunter/TDT. Note that this function may not work under platforms other than linux, and may not work with your version of genehunter. As a matter of fact, it is almost unrelated to simuPOP and is provided only as an example how to use python to analyze data.

Parameters: *file:* file to analyze. This function will look for file.dat and file.pre in linkage format.

loci: a list of loci at which p-value will be returned. If the list is empty, all p-values are returned.

gh: name (or full path) of genehunter executable. Default to 'gh'

Return value: A list (for each chromosome) of list (for each locus) of p-values.

VC_merlin (*file, merlin='merlin'*)
run variance component method

file: file.ped, file.dat, file.map and file,mdl are expected. file can contain directory name.

saveFstat (*output='', outputExpr='', **kwargs*)
operator version of the function SaveFstat

saveLinkage (*output='', outputExpr='', **kwargs*)
An operator to save population in linkage format

3.3.3 Module `simuRPy`

This module helps the use of rpy package with simuPOP. It defines an operator varPlotter that can be used to plot population expressions when rpy is installed.

Module Functions

`rmatrix`(*mat*)

Convert a Python 2d list to r matrix format that can be passed to functions like image directly.

`varPlotter`(*self, expr, history=True, varDim=1, numRep=1, win=0, ylim=[0, 0], update=1, title='', xlab='generation', ylab='', axes=True, lty=[], col=[], mfrow=[1, 1], separate=False, byRep=False, byVal=False, plotType='plot', level=20, saveAs='', leaveOpen=True, dev='', width=0, height=0, *args, **kwargs*)

Plotting with history

plot a number in the form of a variable or expression, use

```
>>> varPlotter(var='expr')
```

plot a vector in the same window and there is only one replicate in the simulator, use

```
>>> varPlotter(var='expr', varDim=len)
```

where len is the dimension of your variable or expression. Each line in the figure represents the history of an item in the array.

plot a vector in the same window and there are several replicates, use

```
>>> varPlotter(var='expr', varDim=len, numRep=nr, byRep=1)
```

varPlotter will try to use an appropriate layout for your subplots (for example, use 3x4 if numRep=10). You can also specify parameter mfrow to change the layout.

if you would like to plot each item of your array variables in a subplot, use

```
>>> varPlotter(var='expr', varDim=len, byVal=1)
```

or in case of a single replicate

```
>>> varPlotter(var='expr', varDim=len, byVal=1, numRep=nr)
```

There will be numRep lines in each subplot.

Use option `history=False` to plot with history. Parameters byVal , varDim etc. will be ignored.

Other options are

title, xtitle, ytitle: title of your figure(s). title is default to your expression, xtitle is defaulted to generation.

win: window of generations. I.e., how many generations to keep in a figure. This is useful when you want to keep track of only recent changes.

update: update figure after update generations. This is used when you do not want to update the figure at every generation.

saveAs: save figures in files saveAs#gen.eps. For example, if saveAs='demo', you will get files demo1.eps, demo2.eps etc.

separate: plot data lines in separate panels.

image: use R image function to plot image, instead of lines.

level: level of image colors (default to 20).

leaveOpen: whether or not leave the plot open when plotting is done. Default to True.

3.3.4 Module `hapMapUtil`

Utility functions to manipulate HapMap data. These functions are provided as examples on how to load and evolve the HapMap dataset. They tend to change frequently so do not call these functions directly. It is recommended that you copy these function to your script when you need to use them.

Module Functions

evolveHapMap(*pop, endingSize, gen, migr=<simuPOP_std.noneOp; proxy of <Swig Object of type 'simuPOP::noneOp *' at 0x44e4830> >, expand='exponential', mergeAt=10000, initMultiple=1, recIntensity=0.01, mutRate=9.9999999999999995e-08, step=10, keepParents=False, numOffspring=1, recordAncestry=False*)

Evolve and expand the hapmap population

gen: total evolution generation

initMultiple: copy each individual initMultiple times, to avoid: rapid loss of genotype variation when population size is small.

endingSize: ending poplation size

expand: expanding method, can be linear or exponential

mergeAt: when to merge population?

gen: generations to evolve

migr: a migrator to be used.

recIntensity: recombination intensity

mutRate: mutation rate

step: step at which to display statistics

keepParents: whether or not keep parental generations

numOffspring: number of offspring at the last generation

recordAncestry: whether or not calculate ancestry to an information field: ancestry. Only usable with two hapmap populations.

getMarkersFromName(*HapMap_dir, names, chroms=[], hapmap_pops=[], minDiffAF=0, numMarkers=[]*)

Get population from marker names. This function: returns a tuple with a population with found markers and names of markers that can not be located in the HapMap data. The returned population has three subpopulations, corresponding to CEU, YRI and JPT+CHB HapMap populations.

HapMap_dir: where HapMap data in simuPOP format is stored. The files: should have been prepared by scripts/loadHapMap.py.

names: names of markers. It can either be a stright list of names, or: a dictionary of names categorized by chromosome number.

chroms: a list of chromosomes to look in. If empty, all 22 autosomes: will be tried. Chromosome index starts from 1. (1, ..., 22).

hapmap_pops: hapmap populations to load, can be a list of 'CEU', 'YRI': or 'JPT+CHB', or a list of 0, 1, 2. If empty (default), all three populations will be loaded.

minDiffAF: minimal allele frequency difference between hapmap populations.: If three subpopulations are loaded, use the maximal of three pair-wise allele frequency differences for comparison. This option is ignored if hapmap_pops has length one.

numMarkers: number of markers to use for each chromosome. Must have: the same length as chroms.

getMarkersFromRange(*HapMap_dir, hapmap_pops, chrom, startPos, endPos, maxNum, minAF=0, minDiffAF=0, minDist=0, maxDist=0*)

Get a population with markers from given range

HapMap_dir: where HapMap data in simuPOP format is stored. The files: should have been prepared by scripts/loadHapMap.py.

hapmap_pops: HapMap populations to load. It can be a list of 'CEU', 'YRI': or 'JPT+CHB', or a list of 0, 1, 2. If empty, all hapmap populations will be loaded.

chrom: chromosome number (1-based index)

startPos: starting position (in cM)

endPos: ending position (in cM). If 0, ignore this parameter.

maxNum: maximum number of markers to get. If 0, ignore this parameter.

minAF: minimal minor allele frequency

minDiffAf: minimal allele frequency between HapMap populations.

minDist: minimal distance between two adjacent markers, in cM

maxDist: maximum distance. If exceed, try to pick up a marker ASAP.

`sample1DSL`(*pop, DSL, DA, pene, name, sampleSize*)

Sample from the final population, using a single lcous penetrance model.

DSL: disease locus

DA: disease allele

pene: penetrance

name: name of directory to save (it must exist)

sampleSize: sample size, in this case, sampleSize/4 is the number of families

`sample2DSL`(*pop, DSL, pene, name, size*)

Sample from the final population, using a two locus penetrance model

DSL: disease loci (two locus)

pene: penetrance value, assuming a two-locus model

name: name to save sample

size: sample size

INDEX