

---

# simuPOP Reference Manual

*Release 0.8.8 (Rev: 2000 )*

Bo Peng

December 2004

Last modified  
November 27, 2008

**Department of Epidemiology, U.T. M.D. Anderson Cancer Center**

**Email:** [bpeng@mdanderson.org](mailto:bpeng@mdanderson.org)

**URL:** <http://simupop.sourceforge.net>

**Mailing List:** [simupop-list@lists.sourceforge.net](mailto:simupop-list@lists.sourceforge.net)

© 2004-2008 Bo Peng

---

Permission is granted to make and distribute verbatim copies of this manual provided the copyright notice and this permission notice are preserved on all copies. Permission is granted to copy and distribute modified versions of this manual under the conditions for verbatim copying, provided also that the sections entitled Copying and GNU General Public License are included exactly as in the original, and provided that the entire resulting derived work is distributed under the terms of a permission notice identical to this one. Permission is granted to copy and distribute translations of this manual into another language, under the above conditions for modified versions, except that this permission notice may be stated in a translation approved by the Free Software Foundation.

## 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.

simuPOP is written in C++ and is provided as Python modules. It 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/Spplus or Matlab-like environment where users can interactively create, manipulate and evolve populations, monitor and visualize population statistics and apply gene mapping methods. Please refer to the *simuPOP user's guide* for a detailed introduction to simuPOP concepts, and a number of examples on how to use simuPOP to perform various simulations.

This reference manual lists all variables, functions, classes and utility modules of simuPOP. Please report any error to the simuPOP mailing list `simupop-list@lists.sourceforge.net`.

### How to cite simuPOP:

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

Bo Peng and Christopher Amos (2008) Forward-time simulations of nonrandom mating populations using simuPOP. *bioinformatics*, **24** (11): 1408-1409.



# CONTENTS

<b>1</b>	<b>simuPOP Components</b>	<b>3</b>
1.1	Individual, population and simulator	3
1.1.1	Class <code>GenoStruTrait</code>	3
1.1.2	Class <code>individual</code>	5
1.1.3	Class <code>population</code>	7
1.1.4	Class <code>simulator</code>	12
1.1.5	Class <code>pedigree</code>	14
1.2	Virtual subpopulation splitters	15
1.2.1	Class <code>vspSplitter</code>	15
1.2.2	Class <code>sexSplitter</code>	15
1.2.3	Class <code>affectionSplitter</code>	15
1.2.4	Class <code>infoSplitter</code>	16
1.2.5	Class <code>proportionSplitter</code>	16
1.2.6	Class <code>rangeSplitter</code>	16
1.2.7	Class <code>genotypeSplitter</code>	17
1.2.8	Class <code>combinedSplitter</code>	17
1.3	Mating Scheme	17
1.3.1	Class <code>mating</code>	17
1.3.2	Class <code>noMating</code>	18
1.3.3	Class <code>cloneMating</code>	19
1.3.4	Class <code>binomialSelection</code>	19
1.3.5	Class <code>baseRandomMating</code>	19
1.3.6	Class <code>randomMating</code>	20
1.3.7	Class <code>selfMating</code>	20
1.3.8	Class <code>monogamousMating</code>	21
1.3.9	Class <code>polygamousMating</code>	21
1.3.10	Class <code>consanguineousMating</code>	21
1.3.11	Class <code>alphaMating</code>	22
1.3.12	Class <code>haplodiploidMating</code>	22
1.3.13	Class <code>pyMating</code>	23
1.3.14	Class <code>heteroMating</code>	23
1.3.15	Class <code>sequentialParentChooser</code>	23
1.3.16	Class <code>sequentialParentsChooser</code>	23
1.3.17	Class <code>randomParentChooser</code>	24
1.3.18	Class <code>randomParentsChooser</code>	24
1.3.19	Class <code>infoParentsChooser</code>	25
1.3.20	Class <code>pyParentsChooser</code>	25
1.3.21	Class <code>cloneOffspringGenerator</code>	25
1.3.22	Class <code>selfingOffspringGenerator</code>	26

1.3.23	Class haplodiploidOffspringGenerator	26
1.3.24	Class mendelianOffspringGenerator	26
<b>2</b>	<b>Operator References</b>	<b>27</b>
2.1	Common interface of all operators	27
2.1.1	Class baseOperator	27
2.2	Initialization	28
2.2.1	Class initializer	28
2.2.2	Class initSex (Function InitSex)	29
2.2.3	Class initByFreq (Function InitByFreq)	29
2.2.4	Class initByValue (Function InitByValue)	29
2.2.5	Class spread (Function Spread)	30
2.2.6	Class pyInit (Function PyInit)	30
2.3	Migration	31
2.3.1	Class migrator	31
2.3.2	Class pyMigrator	32
2.3.3	Class splitSubPop (Function SplitSubPop)	33
2.3.4	Class mergeSubPops (Function MergeSubPops)	33
2.3.5	Class resizeSubPops (Function ResizeSubPops)	33
2.4	Mutation	34
2.4.1	Class mutator	34
2.4.2	Class kamMutator (Function KamMutate)	35
2.4.3	Class smmMutator (Function SmmMutate)	35
2.4.4	Class gsmMutator (Function GsmMutate)	35
2.4.5	Class pyMutator (Function PyMutate)	36
2.4.6	Class pointMutator (Function PointMutate)	36
2.5	Recombination and gene conversion	37
2.5.1	Class recombinator	37
2.6	Selection	38
2.6.1	Class selector	38
2.6.2	Class mapSelector (Function MapSelector)	39
2.6.3	Class maSelector (Function MaSelect)	40
2.6.4	Class mlSelector (Function MlSelect)	40
2.6.5	Class pySelector (Function PySelect)	41
2.7	Penetrance	41
2.7.1	Class penetrance	41
2.7.2	Class mapPenetrance (Function MapPenetrance)	42
2.7.3	Class maPenetrance (Function MaPenetrance)	42
2.7.4	Class mlPenetrance (Function MlPenetrance)	43
2.7.5	Class pyPenetrance (Function PyPenetrance)	43
2.8	Quantitative Trait	44
2.8.1	Class quanTrait	44
2.8.2	Class mapQuanTrait (Function MapQuanTrait)	44
2.8.3	Class maQuanTrait (Function MaQuanTrait)	45
2.8.4	Class mlQuanTrait (Function MlQuanTrait)	45
2.8.5	Class pyQuanTrait (Function PyQuanTrait)	46
2.9	Ascertainment	46
2.10	Statistics Calculation	46
2.10.1	Class stator	46
2.10.2	Class stat (Function Stat)	46
2.11	Expression and Statements	50
2.11.1	Class dumper	50
2.11.2	Class savePopulation	51
2.11.3	Class pyOutput	51

2.11.4	Class <code>pyEval</code> (Function <code>PyEval</code> )	52
2.11.5	Class <code>pyExec</code> (Function <code>PyExec</code> )	52
2.11.6	Class <code>infoEval</code> (Function <code>infoEval</code> )	52
2.11.7	Class <code>infoExec</code> (Function <code>infoExec</code> )	53
2.12	Tagging (used for pedigree tracking)	54
2.12.1	Class <code>tagger</code>	54
2.12.2	Class <code>inheritTagger</code>	54
2.12.3	Class <code>parentTagger</code>	54
2.12.4	Class <code>parentsTagger</code>	55
2.12.5	Class <code>sexTagger</code>	55
2.12.6	Class <code>affectionTagger</code>	55
2.12.7	Class <code>infoTagger</code>	56
2.12.8	Class <code>pyTagger</code>	56
2.13	Terminator	56
2.13.1	Class <code>terminateIf</code>	56
2.14	Python operators	57
2.14.1	Class <code>pyOperator</code>	57
2.14.2	Class <code>pyIndOperator</code>	58
2.15	Miscellaneous	58
2.15.1	Class <code>ifElse</code>	58
2.15.2	Class <code>turnOnDebug</code> (Function <code>TurnOnDebug</code> )	59
2.15.3	Class <code>turnOffDebug</code> (Function <code>TurnOffDebug</code> )	59
2.15.4	Class <code>noneOp</code>	59
2.15.5	Class <code>pause</code>	60
2.15.6	Class <code>ticToc</code> (Function <code>TicToc</code> )	60
2.15.7	Class <code>setAncestralDepth</code>	60
<b>3</b>	<b>Global and Python Utility functions</b>	<b>61</b>
3.1	Global functions	61
3.2	Utility Classes	62
3.2.1	Class <code>RNG</code>	62
3.3	Module <code>simuOpt</code>	63
3.3.1	<code>simuOpt</code> functions	63
3.4	Module <code>simuUtil</code>	66
3.4.1	Class <code>randomSample</code>	66
3.4.2	Class <code>caseControlSample</code>	66
3.4.3	Class <code>affectedSibpairSample</code>	67
3.4.4	Class <code>pySubset</code>	67
3.4.5	Class <code>simuProgress</code>	67
3.4.6	<code>simuUtil</code> functions	68
3.5	Module <code>simuRPy</code>	74
3.5.1	Class <code>varPlotter</code>	74
<b>Index</b>		<b>77</b>









# simuPOP Components

## 1.1 Individual, population and simulator

### 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 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.

#### **`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 `#-ploidy` where # is the ploidy number.

#### **`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 list of the names of all chromosomes.

**chromType** (*chrom*)  
Return the type of a chromosome *chrom* (1 for Autosome, 2 for ChromosomeX, 3 for ChromosomeY, and 4 for Mitochondrial).

**chromTypes** ()  
Return the type of all chromosomes (1 for Autosome, 2 for ChromosomeX, 3 for ChromosomeY, and 4 for Mitochondrial).

**numChrom** ()  
Return the number of chromosomes.

**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*.

**lociByName** (*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* parameter 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.

**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. An `IndexError` will be raised if *allele* is larger than the maximum allowed allele state of this module (`MaxAllele()`).

**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 so `alleleNames() [allele]` might fail (use `alleleNames(allele)` instead).

**infoField** (*idx*)

Return the name of information field *idx*.

**infoFields** ()

Return a list 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.

### 1.1.2 Class individual

A population consists of individuals with the same genotypic structure. An individual object cannot be created independently, but references to individuals can be retrieved using member functions of a population object. In addition to structural information shared by all individuals in a population (provided by class `genoStruTrait`), an `individual` class provides member functions to get and set *genotype*, *sex*, *affection status* and *information fields* of an individual.

Genotypes of an individual are stored sequentially and can be accessed locus by locus, or in batch. The alleles are arranged by position, chromosome and ploidy. That is to say, the first allele on the first chromosome of the first homologous set is followed by alleles at other loci on the same chromosome, then markers on the second and later chromosomes, followed by alleles on the second homologous set of the chromosomes for a diploid individual. A consequence of this memory layout is that alleles at the same locus of a non-haploid individual are separated by `individual::totNumLoci()` loci. It is worth noting that access to invalid chromosomes, such as the Y chromosomes of female individuals, are not restricted.

**class individual** ()

An individual object cannot be created directly. It has to be accessed from a population object using functions such as `population::individual(idx)`.

**allele** (*idx*)

Return the current allele at a locus, using its absolute index *idx*.

**allele** (*idx*, *p*)

Return the current allele at locus *idx* on the *p*-th set of homologous chromosomes.

**allele** (*idx*, *p*, *chrom*)

Return the current allele at locus *idx* on chromosome *chrom* of the *p*-th set of homologous chromosomes.

**setAllele** (*allele*, *idx*)

Set allele *allele* to a locus, using its absolute index *idx*.

**setAllele** (*allele*, *idx*, *p*)

Set allele *allele* to locus *idx* on the *p*-th homologous set of chromosomes.

**setAllele** (*allele*, *idx*, *p*, *chrom*)

Set allele *allele* to locus *idx* on chromosome *chrom* of the *p*-th homologous set of chromosomes.

**genotype()**  
Return an editable array (a carrary of length `totNumLoci()` \* `ploidy()`) that represents all alleles of an individual.

**genotype(*p*)**  
Return an editable array (a carrary of length `totNumLoci()`) that represents all alleles on the *p*-th homologous set of chromosomes.

**genotype(*p*, *chrom*)**  
Return an editable array (a carrary of legnth `numLoci(chrom)`) that represents all alleles on chromosome *chrom* of the *p*-th homologous set of chromosomes.

**setGenotype(*geno*)**  
Fill the genotype of an individual using a list of alleles *geno*. *geno* will be reused if its length is less than `totNumLoci()` \* `ploidy()`.

**setGenotype(*geno*, *p*)**  
Fill the genotype of the *p*-th homologous set of chromosomes using a list of alleles *geno*. *geno* will be reused if its length is less than `totNumLoci()`.

**setGenotype(*geno*, *p*, *chrom*)**  
Fill the genotype of chromosome *chrom* on the *p*-th homologous set of chromosomes using a list of alleles *geno*. *geno* will be reused if its length is less than `numLoci(chrom)`.

**setSex(*sex*)**  
Set individual sex to Male or Female.

**sex()**  
Return the sex of an individual, 1 for male and 2 for female.

**sexChar()**  
Return the sex of an individual, M for male or F for female.

**affected()**  
Return `True` if this individual is affected.

**affectedChar()**  
Return A if this individual is affected, or U otherwise.

**setAffected(*affected*)**  
Set affection status to *affected* (`True` or `False`).

**info(*idx*)**  
Return the value of an information field *idx* (an index).

**info(*name*)**  
Return the value of an information field *name*.

**intInfo(*idx*)**  
Return the value of an information field *idx* (an index) as an integer number.

**intInfo(*name*)**  
Return the value of an information field *name* as an integer number.

**setInfo(*value*, *idx*)**  
Set the value of an information field *idx* (an index) to *value*.

**setInfo(*value*, *name*)**  
Set the value of an information field *name* to *value*.

### 1.1.3 Class `population`

A `simuPOP` population consists of individuals of the same genotypic structure, organized by generations, subpopulations and virtual subpopulations. It also contains a Python dictionary that is used to store arbitrary population variables. In addition to genotypic structured related functions provided by the `genoStruTrait` class, the `population` class provides a large number of member functions that can be used to

- Create, copy and compare populations.
- Manipulate subpopulations. A population can be divided into several subpopulations. Because individuals only mate with individuals within the same subpopulation, exchange of genetic information across subpopulations can only be done through migration. A number of functions are provided to access subpopulation structure information, and to merge and split subpopulations.
- Define and access virtual subpopulations. A *virtual subpopulation splitter* can be assigned to a population, which defines groups of individuals called *virtual subpopulations* (VSP) within each subpopulation.
- Access individuals individually, or through iterators that iterate through individuals in (virtual) subpopulations.
- Access genotype and information fields of individuals at the population level. From a population point of view, all genotypes are arranged sequentially individual by individual. Please refer to class `individual` for an introduction to genotype arrangement of each individual.
- Store and access *ancestral generations*. A population can save arbitrary number of ancestral generations. It is possible to directly access an ancestor, or make an ancestral generation the current generation for more efficient access.
- Insert or remove loci, resize (shrink or expand) a population, sample from a population, or merge with other populations.
- Manipulate population variables and evaluate expressions in this *local namespace*.
- Save and load a population.

**class `population`** (`size=[]`, `ploidy=2`, `loci=[]`, `chromTypes=[]`, `lociPos=[]`, `ancestralGens=0`, `chromNames=[]`, `alleleNames=[]`, `lociNames=[]`, `infoFields=[]`)

The following parameters are used to create a population object:

*size*: A list of subpopulation sizes. The length of this list determines the number of subpopulations of this population. If there is no subpopulation, `size=[popSize]` can be written as `size=popSize`.

*ploidy*: Number of homologous sets of chromosomes. Default to 2 (diploid). For efficiency considerations, all chromosomes have the same number of homologous sets, even if some chromosomes (e.g. mitochondrial) or some individuals (e.g. males in a haplodiploid population) have different numbers of homologous sets. The first case is handled by setting *chromTypes* of each chromosome. Only the haplodiploid populations are handled for the second case, for which `ploidy=Haplodiploid` should be used.

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

*chromTypes*: A list that specifies the type of each chromosome, which can be `Autosome`, `ChromosomeX`, `ChromosomeY`, or `Mitochondrial`. All chromosomes are assumed to be autosomes if this parameter is ignored. Sex chromosome can only be specified in a diploid population where the sex of an individual is determined by the existence of these chromosomes using the XX (Female) and XY (Male) convention. Both sex chromosomes have to be available and be specified only once. Because chromosomes X and Y are treated as two chromosomes, recombination on the pseudo-autosomal regions of the sex chromosomes is not supported. A `Mitochondrial` chromosome only exists in females and is inherited maternally.

*lociPos*: Positions of all loci on all chromosome, as a list of float numbers. Default to 1, 2, ... etc on each chromosome. Positions on the same chromosome should be ordered. A nested list that specifies positions of loci on each chromosome is also acceptable.

*ancestralGens*: Number of the most recent ancestral generations to keep during evolution. Default to 0, which means only the current generation will be kept. If it is set to -1, all ancestral generations will be kept in this population (and exhaust your computer RAM quickly).

*chromNames*: A list of chromosome names. Default to `chrom1`, `chrom2`, ... etc.

*alleleNames*: A list of allele names for all markers. For example, *alleleNames*=( 'A' , 'C' , 'T' , 'G' ) names allele 0 – 3 'A' , 'C' , 'T' , and 'G' respectively. Note that simuPOP does not yet support locus-specific allele names.

*lociNames*: A list or a matrix (separated by chromosomes) of names for each locus. Default to "locX-Y" where X and Y are 1-based chromosome and locus indexes, respectively.

*infoFields*: Names of information fields (named float number) that will be attached to each individual.

**ancestor** (*ind*, *gen*)

Reference to an individual *ind* in an ancestral generation

**ancestor** (*ind*, *subPop*, *gen*)

Reference to an individual *ind* in a specified subpopulation or an ancestral generation

**clone** ()

Create a cloned copy of a population. Note that Python statement `pop1 = pop` only creates a reference to an existing population *pop*.

**save** (*filename*)

Save population to a file *filename*, which can be loaded by a global function `LoadPopulation(filename)`.

**absIndIndex** (*idx*, *subPop*)

Return the absolute index of an individual *idx* in subpopulation *subPop*.

**numSubPop** ()

Return the number of subpopulations in a population. Return 1 if there is no subpopulation structure.

**popSize** ()

Return the total number of individuals in all subpopulations.

**subPopBegin** (*subPop*)

Return the index of the first individual in subpopulation *subPop*. An `IndexError` will be raised if *subPop* is out of range.

**subPopEnd** (*subPop*)

Return the index of the last individual in subpopulation *subPop* plus 1, so that `range(subPopBegin(subPop), subPopEnd(subPop))` can iterate through the index of all individuals in subpopulation *subPop*.

**subPopIndPair** (*idx*)

Return the subpopulation ID and relative index of an individual, given its absolute index *idx*.

**subPopSize** (*subPop*)

Return the size of a subpopulation (`subPopSize(sp)`) or a virtual subpopulation (`subPopSize([sp, vsp])`).

**subPopSizes** ()

Return the sizes of all subpopulations in a list. Virtual subpopulations are not considered.

**numVirtualSubPop** ()

Return the number of virtual subpopulations (VSP) defined by a VSP splitter. Return 0 if no VSP is defined.



**setVirtualSplitter** (*splitter*)

Set a VSP *splitter* to the population, which defines the same VSPs for all subpopulations. If different VSPs are needed for different subpopulations, a `combinedSplitter` can be used to make these VSPs available to all subpopulations.

**virtualSubPopName** (*subPop*)

Return the name of a virtual subpopulation *subPop* (specified by a (*sp*, *vsp*) pair). Because VSP names are the same across all subpopulations, a single VSP index is also acceptable.

**individual** (*idx*)

Return a reference to individual *ind* in the population.

**individual** (*idx*, *subPop*)

Return a reference to individual *ind* in subpopulation *subPop*.

**individuals** ()

Return a Python iterator that can be used to iterate through all individuals in a population.

**individuals** (*subPop*)

Return an iterator that can be used to iterate through all individuals in a subpopulation (*subPop*=*spID*) or a virtual subpopulation (*subPop*=[*spID*, *vspID*]).

**genotype** ()

Return an editable array of the genotype of all individuals in this population.

**genotype** (*subPop*)

Return an editable array of the genotype of all individuals in subpopulation *subPop*.

**setGenotype** (*geno*)

Fill the genotype of all individuals of a population using a list of alleles *geno*. *geno* will be reused if its length is less than `popSize()*totNumLoci()*ploidy()`.

**setGenotype** (*geno*, *subPop*)

Fill the genotype of all individuals of in subpopulation *subPop* using a list of alleles *geno*. *geno* will be reused if its length is less than `subPopSize(subPop)*totNumLoci()*ploidy()`.

**ancestor** (*idx*, *gen*)

Return a reference to individual *idx* in ancestral generation *gen*. The correct individual will be returned even if the current generation is not the present one (see also `useAncestralGen`).

**ancestor** (*ind*, *subPop*, *gen*)

Return a reference to individual *idx* of subpopulation *subPop* in ancestral generation *gen*.

**ancestralGens** ()

Return the actual number of ancestral generations stored in a population, which does not necessarily equal to the number set by `setAncestralDepth()`.

**push** (*pop*)

Push population *pop* into the current population. Both populations should have the same genotypic structure. The current population is discarded if *ancestralDepth* (maximum number of ancestral generations to hold) is zero so no ancestral generation can be kept. Otherwise, the current population will become the parental generation of *pop*, advancing the greatness level of all existing ancestral generations by one. If *ancestralDepth* is positive and there are already *ancestralDepth* ancestral generations (see also: `ancestralGens()`), the greatest ancestral generation will be discarded. In any case, population *pop* becomes invalid as all its individuals are absorbed by the current population.

**setAncestralDepth** (*depth*)

Set the intended ancestral depth of a population to *depth*, which can be 0 (does not store any ancestral generation), -1 (store all ancestral generations), and a positive number (store *depth* ancestral generations).

**useAncestralGen** (*idx*)

Making ancestral generation *idx* (0 for current generation, 1 for parental generation, 2 for grand-parental

generation, etc) the current generation. This is an efficient way to access population properties of an ancestral generation. `useAncestralGen(0)` should always be called afterward to restore the correct order of ancestral generations.

**addChrom** (*lociPos*, *lociNames*=[], *chromName*="", *chromType*=Autosome)

Add chromosome *chromName* with given type *chromType* to a population, with loci *lociNames* inserted at position *lociPos*. *lociPos* should be ordered. *lociNames* and *chromName* should not exist in the current population. If they are not specified, simuPOP will try to assign default names, and raise a `ValueError` if the default names have been used.

**addChromFromPop** (*pop*)

Add chromosomes in population *pop* to the current population. Population *pop* should have the same number of individuals as the current population in the current and all ancestral generations. This function merges genotypes on the new chromosomes from population *pop* individual by individual.

**addIndFromPop** (*pop*)

Add all individuals, including ancestors, in *pop* to the current population. Two populations should have the same genotypic structures and number of ancestral generations. Subpopulations in population *pop* are kept.

**addLoci** (*chrom*, *pos*, *names*=[])

Insert loci *names* at positions *pos* on chromosome *chrom*. These parameters should be lists of the same length, although *names* may be ignored, in which case random names will be given. Alleles at inserted loci are initialized with zero alleles. Note that loci have to be added to existing chromosomes. If loci on a new chromosome need to be added, function `addChrom` should be used. This function returns indexes of the inserted loci.

**addLociFromPop** (*pop*)

Add loci from population *pop*, chromosome by chromosome. Added loci will be inserted according to their position. Their position and names should not overlap with any locus in the current population. Population *pop* should have the same number of individuals as the current population in the current and all ancestral generations.

**extract** (*field*=None, *loci*=None, *infoFields*=None, *ancGen*=-1)

Extract subsets of individuals, loci and/or information fields from the current population and create a new one. If information field *field* is not None, individuals with negative values at this information field will be removed, and others are put into subpopulations specified by this field. If *loci* is not None, only genotypes at *loci* are extracted. If *infoFields* is not None, only these information fields will be extracted. If *ancGen* is not -1 (default, meaning all ancestral generations), only *ancGen* ancestral generations will be kept. As an advanced feature, *field* can be information field of a pedigree object *ped*. This allows extraction of individuals according to pedigrees identified in a pedigree object. This pedigree should have the same number of individuals in all generations.

**mergeSubPops** (*subPops*=[])

Merge subpopulations *subPops*. If *subPops* is empty (default), all subpopulations will be merged. *subPops* do not have to be adjacent to each other. They will all be merged to the subpopulation with the smallest subpopulation ID. Indexes of the rest of the subpopulation may be changed.

**removeEmptySubPops** ()

Remove empty subpopulations by adjusting subpopulation IDs.

**removeIndividuals** (*inds*)

Remove individuals *inds* (absolute indexes) from the current population. A subpopulation will be kept even if all individuals from it are removed. This function only affects the current generation.

**removeLoci** (*loci*=[], *keep*=[])

Remove *loci* (absolute indexes) and genotypes at these loci from the current population. Alternatively, a parameter *keep* can be used to specify loci that will not be removed.

**removeSubPops** (*subPops*)

Remove subpopulations *subPop* and all their individuals. Indexes of subpopulations after removed subpopulations will be shifted.

**resize** (*newSubPopSizes*, *propagate=False*)

Resize population by giving new subpopulation sizes *newSubPopSizes*. Individuals at the end of some subpopulations will be removed if the new subpopulation size is smaller than the old one. New individuals will be appended to a subpopulation if the new size is larger. Their genotypes will be set to zero (default), or be copied from existing individuals if *propagate* is set to `True`. More specifically, if a subpopulation with 3 individuals is expanded to 7, the added individuals will copy genotypes from individual 1, 2, 3, and 1 respectively. Note that this function only resizes the current generation. This function affects only the current generation.

**setSubPopByIndInfo** (*field*)

Rearrange individuals to their new subpopulations according to their integer values at information field *field* (value returned by `individual::indInfo(field)`). Individuals with negative values at this *field* will be removed.

**splitSubPop** (*subPop*, *sizes*)

Split subpopulation *subPop* into subpopulations of given *sizes*, which should add up to the size of subpopulation *subPop*. Alternatively, *sizes* can be a list of proportions (add up to 1) from which the sizes of new subpopulations are determined. If *subPop* is not the last subpopulation, subpopulation indexes will be changed. For example, if you split the second subpopulation into two, a population with three subpopulations will have four subpopulations: 0 (untouched), 1.1→1, 1.2→2, 2→3 (changed).

**addInfoField** (*field*, *init=0*)

Add an information field *field* to a population and initialize its values to *init*.

**addInfoFields** (*fields*, *init=0*)

Add a list of information fields *fields* to a population and initialize their values to *init*. If an information field already exists, it will be re-initialized.

**indInfo** (*idx*)

Return the information field *idx* (an index) of all individuals as a list.

**indInfo** (*name*)

Return the information field *name* of all individuals as a list.

**indInfo** (*idx*, *subPop*)

Return the information field *idx* (an index) of all individuals in (virtual) subpopulation *subPop* as a list.

**indInfo** (*name*, *subPop*)

Return the information field *name* of all individuals in (virtual) subpopulation *subPop* as a list.

**setIndInfo** (*values*, *idx*)

Set information field *idx* (an index) of the current population to *values*. *values* will be reused if its length is smaller than `popSize()`.

**setIndInfo** (*values*, *name*)

Set information field *name* of the current population to *values*. *values* will be reused if its length is smaller than `popSize()`.

**setIndInfo** (*values*, *idx*, *subPop*)

Set information field *idx* (an index) of a subpopulation (`subPop=sp`) or a virtual subpopulation (`subPop=[sp, vsp]`) to *values*. *values* will be reused if its length is smaller than `subPopSize(subPop)`.

**setIndInfo** (*values*, *name*, *subPop*)

Set information field *name* of a subpopulation (`subPop=sp`) or a virtual subpopulation (`subPop=[sp, vsp]`) to *values*. *values* will be reused if its length is smaller than `subPopSize(subPop)`.

**setInfoFields** (*fields*, *init*=0)

Set information fields *fields* to a population and initialize them with value *init*. All existing information fields will be removed.

**vars** ()

Return variables of a population as a Python dictionary.

**vars** (*subPop*)

Return a dictionary `vars () ["subPop"] [subPop]`. *subPop* can be a number (`subPop=spID`), or a pair of numbers (`subPop=(spID, vspID)`). A `ValueError` will be raised if key 'subPop' does not exist in `vars ()`, or if key *subPop* does not exist in `vars () ["subPop"]`.

**dvars** ()

Return a wrapper of Python dictionary returned by `vars()` so that dictionary keys can be accessed as attributes. For example `pop.dvars().alleleFreq` is equivalent to `pop.vars () ["alleleFreq"]`.

**dvars** (*subPop*)

Return a wrapper of Python dictionary returned by `vars (subPop)` so that dictionary keys can be accessed as attributes.

#### 1.1.4 Class simulator

A `simuPOP` simulator is responsible for evolving one or more replicates of a *population* forward in time, subject to various *operators*. Populations in a simulator are created as identical copies of a population and will become different after evolution. A *mating scheme* needs to be specified, which will be used to generate offspring generations during evolution. A number of functions are provided to access simulator properties, access populations and their variables, copy, save and load a simulator.

The most important member function of a simulator is `evolve`, which evolves populations forward in time, subject to various *operators*. Because populations in a simulator have to keep the same genotypic structure, several functions are provided to change ancestral depth and information fields of all populations. These functions cannot be replaced by similar calls to all populations in a simulator because the genotypic structure of the simulator itself needs to be updated.

**class simulator** (*pop*, *matingScheme*, *rep*=1)

Create a simulator with *rep* replicates of population *pop*. Population *pop* will be copied *rep* times (default to 1), while keeping the passed population intact. A mating scheme *matingScheme* will be used to evolve these populations.

**clone** ()

Clone a simulator, along with all its populations. Note that Python assign statement `simul = simu` only creates a symbolic link to an existing simulator.

**save** (*filename*)

Save a simulator to file *filename*, which can be loaded by a global function `LoadSimulator`.

**gen** ()

Return the current generation number, which is the initial generation number (0, or some value set by `setGen (gen)`) plus the total number of generations evolved.

**setGen** (*gen*)

Set the current generation number of a simulator to *gen*.

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

Evolve all populations *gen* generations, subject to operators *ops*, *preOps* and *postOps*. Operators *preOps*

are applied to all populations (subject to applicability restrictions of the operators, imposed by the *rep* parameter of these operators) before evolution. They are usually used to initialize populations. Operators *postOps* are applied to all populations after the evolution.

Operators *ops* are applied during the life cycle of each generation. Depending on the stage of these operators, they can be applied before-, during-, and/or post-mating. These operators can be applied at all or some of the generations, depending the *begin*, *end*, *step*, and *at* parameters of these operators. Populations in a simulator are evolved one by one. At each generation, the applicability of these operators are determined. Pre-mating operators are applied to a population first. A mating scheme is then used to populate an offspring generation, using applicable during-mating operators. After an offspring generation is successfully generated and becomes the current generation, applicable post-mating operators are applied to it. Because the order at which operators are applied can be important, and the stage(s) at which operators are applied are not always clear, a parameter *dryRun* can be used. If set to `True`, this function will print out the order at which all operators are applied, without actually evolving the populations.

Parameter *gen* can be set to a positive number, which is the number of generations to evolve. If *gen* is negative (default), the evolution will continue indefinitely, until all replicates are stopped by a special kind of operators called *terminators*. At the end of the evolution, the generations that each replicates have evolved are returned.

**extract** (*rep*)

Extract the *rep-th* population from a simulator. This will reduce the number of populations in this simulator by one.

**numRep** ()

Return the number of replicates.

**population** (*rep*, *byRef=True*)

Return the *rep-th* population of a simulator, in the form of a reference (*byRef=True* (default)), or a cloned copy. In the first case, a temporary reference is returned, which will become invalid once the simulator starts evolving or becomes invalid (removed). Modifying the returned object is discouraged because it will change the population within the simulator. In the second case, an independent, cloned copy of the internal population is returned. Modifying the returned population will not change the simulator.

**populations** ()

Return a Python iterator that can be used to iterate through all populations in a simulator.

**addInfoField** (*field*, *init=0*)

Add an information field *field* to all populations in a simulator, and update the genotypic structure of the simulator itself. The information field will be initialized by value *init*.

**addInfoFields** (*fields*, *init=0*)

Add information fields *fields* to all populations in a simulator, and update the genotypic structure of the simulator itself. The information field will be initialized by value *init*.

**setAncestralDepth** (*depth*)

Set ancestral depth of all populations in a simulator.

**setMatingScheme** (*matingScheme*)

Set a new mating scheme *matingScheme* to a simulator.

**vars** (*rep*)

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

**vars** (*rep*, *subPop*)

Return a dictionary of subpopulation variables in a local namespace of the *rep-th* population, equivalent to `x.population(rep).vars(subPop)`.

**dvars** (*rep*)

Return a wrapper of Python dictionary returned by `vars(rep)` so that dictionary keys can

be accessed as attributes. For example `simu.dvars(1).alleleFreq` is equivalent to `simu.vars(1) ["alleleFreq"]`.

**dvars** (*rep, subPop*)

Return a wrapper of Python dictionary returned by `vars(rep, subPop)` so that dictionary keys can be accessed as attributes.

### 1.1.5 Class pedigree

FIXME: No document

**class pedigree** (*pop, loci=[], infoFields=[], ancGen=-1*)

Create a pedigree object from a population, using a subset of loci, information fields and ancestral generations.

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

This function locates relatives (of type *relType*, and sex *relSex*) of each individual and store their indexes in specified information fields *relFields*. The indexes of parents in the parental generation should be available in information fields *parentFields* (default to `['father_idx', 'mother_idx']` which are the information fields used by operator `parentsTagger`. This function currently only work for diploid populations.

*relType*: Relative type, which can be

- REL\_Self set indexes of individual themselves.
- REL\_Spouse locate spouses of individuals in the current generation. A 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

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

*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.

*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).

**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 information 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` individuals referred by *off1* and from generation 0 (*pathGen*[3]) 7. Save index of these individuals to information fields *cousin1* and *cousin2* at generation *pathGen*[0]. In short, this function locates father or mother's brother's daughters.

*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*.

*pathFields*: A list of list of information fields forming a path to trace a certain type of relative.  
*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.  
*pathSex*: (Optional) A list of sex choices, AnySex, Male, Female or OppositeSex, that is used to choose individuals at each step. Default to AnySex.

## 1.2 Virtual subpopulation splitters

### 1.2.1 Class `vspSplitter`

This class is the base class of all virtual subpopulation (VSP) splitters, which provide ways to define groups of individuals in a subpopulation who share certain properties. A splitter defines a fixed number of named VSPs. They do not have to add up to the whole subpopulation, nor do they have to be distinct. After a splitter is assigned to a population, many functions and operators can be applied to individuals within specified VSPs. Only one VSP splitter can be assigned to a population, which defined VSPs for all its subpopulations. If different splitters are needed for different subpopulations, a `combinedSplitter` should be.

**class `vspSplitter()`**

This is a virtual class that cannot be instantiated.

**`clone()`**

All VSP splitter defines a `clone()` function to create an identical copy of itself.

**`name(vsp)`**

Return the name of VSP *vsp* (an index between 0 and `numVirtualSubPop()`).

**`numVirtualSubPop()`**

Return the number of VSPs defined by this splitter.

### 1.2.2 Class `sexSplitter`

This splitter defines two VSPs by individual sex. The first VSP consists of all male individuals and the second VSP consists of all females in a subpopulation.

**class `sexSplitter()`**

Create a sex splitter that defines male and female VSPs.

**`name(vsp)`**

Return "Male" if *vsp*=0 and "Female" otherwise.

**`numVirtualSubPop()`**

Return 2.

### 1.2.3 Class `affectionSplitter`

This class defines two VSPs according individual affection status. The first VSP consists of unaffected individuals and the second VSP consists of affected ones.

**class `affectionSplitter()`**

Create a splitter that defined two VSPs by affection status.

**`name(vsp)`**

Return "Unaffected" if *vsp*=0 and "Affected" if *vsp*=1.

```

numVirtualSubPop()
    Return 2.

```

## 1.2.4 Class infoSplitter

This splitter defines VSPs according to the value of an information field of each individual. A VSP is defined either by a value or a range of values.

```

class infoSplitter (field, values=[], cutoff=[])

```

Create an information splitter using information field *field*. If parameter *values* is specified, each item in this list defines a VSP in which all individuals have this value at information field *field*. If a set of cutoff values are defined in parameter *cutoff*, individuals are grouped by intervals defined by these cutoff values. For example, *cutoff*=[1,2] defines three VSPs with  $v < 1$ ,  $1 \leq v < 2$  and  $v \geq 2$  where  $v$  is the value of an individual at information field *field*. Of course, only one of the parameters *values* and *cutoff* should be defined, values in *cutoff* should be distinct, and in an increasing order.

```

    name (vsp)

```

Return the name of a VSP *vsp*, which is *field* = *value* if VSPs are defined by values in parameter *values*, or *field* < *value* (the first VSP),  $v_1 \leq \text{field} < v_2$  and *field*  $\geq v$  (the last VSP) if VSPs are defined by cutoff values.

```

    numVirtualSubPop()

```

Return the number of VSPs defined by this splitter, which is the length parameter *values* or the length of *cutoff* plus one, depending on which parameter is specified.

## 1.2.5 Class proportionSplitter

This splitter divides subpopulations into several VSPs by proportion.

```

class proportionSplitter (proportions=[])

```

Create a splitter that divides subpopulations by *proportions*, which should be a list of float numbers (between 0 and 1) that add up to 1.

```

    name (vsp)

```

Return the name of VSP *vsp*, which is "Prop *p*" where  $p = \text{proportions}[\text{vsp}]$ .

```

    numVirtualSubPop()

```

Return the number of VSPs defined by this splitter, which is the length of parameter *proportions*.

## 1.2.6 Class rangeSplitter

This class defines a splitter that groups individuals in certain ranges into VSPs.

```

class rangeSplitter (ranges)

```

Create a splitter according to a number of individual ranges defined in *ranges*. For example, *rangeSplitter*(*ranges*=[[0, 20], [40, 50]]) defines two VSPs. The first VSP consists of individuals 0, 1, ..., 19, and the second VSP consists of individuals 40, 41, ..., 49. Note that a nested list has to be used even if only one range is defined.

```

    name (vsp)

```

Return the name of VSP *vsp*, which is "Range [*a*, *b*]" where [*a*, *b*] is range *ranges*[*vsp*].

```

    numVirtualSubPop()

```

Return the number of VSPs, which is the number of ranges defined in parameter *ranges*.



## 1.2.7 Class `genotypeSplitter`

This class defines a VSP splitter that defines VSPs according to individual genotype at specified loci.

**class `genotypeSplitter`** (*loci (or locus)*, *alleles*, *phase=False*)

Create a splitter that defined VSPs by individual genotype at loci *loci* (or *locus* if only one locus is used). Each list in a list *allele* defines a VSP, which is a list of allowed alleles at these *loci*. If only one VSP is defined, the outer list of the nested list can be ignored. If *phase* is true, the order of alleles in each list is significant. If more than one set of alleles are given, individuals having either of them is qualified.

For example, in a haploid population, *locus=1*, *alleles=[0, 1]* defines a VSP with individuals having allele 0 or 1 at locus 1, *alleles=[[0, 1], [2]]* defines two VSPs with individuals in the second VSP having allele 2 at locus 1. If multiple loci are involved, alleles at each locus need to be defined. For example, VSP defined by *loci=[0, 1]*, *alleles=[0, 1, 1, 1]* consists of individuals having alleles [0, 1] or [1, 1] at loci [0, 1].

In a haploid population, *locus=1*, *alleles=[0, 1]* defines a VSP with individuals having genotype [0, 1] or [1, 0] at locus 1. *alleles=[[0, 1], [2, 2]]* defines two VSPs with individuals in the second VSP having genotype [2, 2] at locus 1. If *phase* is set to `True`, the first VSP will only has individuals with genotype [0, 1]. In the multiple loci case, alleles should be arranged by haplotypes, for example, *loci=[0, 1]*, *alleles=[0, 0, 1, 1]*, *phase=True* defines a VSP with individuals having genotype -0-0-, -1-1- at loci 0 and 1. If *phase=False* (default), genotypes -1-1-, -0-0-, -0-1- and -1-0- are all allowed.

**name** (*vsp*)

Return name of VSP *vsp*, which is "Genotype loc1,loc2:genotype" as defined by parameters *loci* and *alleles*.

**numVirtualSubPop** ()

Number of virtual subpops of subpopulation *sp*

## 1.2.8 Class `combinedSplitter`

This splitter takes several splitters and stacks their VSPs together. For example, if the first splitter defines 3 VSPs and the second splitter defines 2, the two VSPs from the second splitter becomes the fourth (index 3) and the fifth (index 4) VSPs of the combined splitter. This splitter is usually used to define different types of VSPs to a population.

**class `combinedSplitter`** (*splitters=[]*)

Create a combined splitter using a list of *splitters*. For example, `combinedSplitter([sexSplitter(), affectionSplitter()])` defines a combined splitter with four VSPs.

**name** (*vsp*)

Return the name of a VSP *vsp*, which is the name a VSP defined by one of the combined splitters.

**numVirtualSubPop** ()

Return the number of VSPs defined by this splitter, which is the sum of the number of VSPs of all combined splitters.

## 1.3 Mating Scheme

### 1.3.1 Class `mating`

The base class of all mating schemes - a required parameter of `simulatorMating` 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*=[], *weight*=0)

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.

*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 determine 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 virtual subpopulation has non-zero weight, this virtual subpopulation will produce no offspring (weight 0).
- any negative number -n: the size will be  $n \cdot m$  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.

**clone** ()

Deep copy of a mating scheme

**submitScratch** (*pop*, *scratch*)

A common submit procedure is defined.

### 1.3.2 Class `noMating`

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*=[], *weight*=0)

Create a scheme with no mating

**Note** All parameters are ignored!

**clone** ()

Deep copy of a scheme with no mating

### 1.3.3 Class `cloneMating`

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, newSub-  
                  PopSize=[], newSubPopSizeExpr="", newSubPopSizeFunc=None, subPop=[], 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
```

### 1.3.4 Class `binomialSelection`

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, new-  
                        SubPopSize=[], newSubPopSizeExpr="", newSubPopSizeFunc=None, subPop=[],  
                        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
```

### 1.3.5 Class `baseRandomMating`

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, polygamous 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`, `monogamousMating`, `polygamousMating`, `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=[], weight=0)
```

FIXME: No document

*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.

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

*polyNum*: Number of sex partners.

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

*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`).

*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`.

**clone ()**

Deep copy of a random mating scheme

### 1.3.6 Class `randomMating`

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=[], weight=0)
```

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

*contWhenUniSex*: Continue when there is only one sex in the population. Default to `True`.

**clone ()**

Deep copy of a random mating scheme

### 1.3.7 Class `selfMating`

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=[], weight=0)
```

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

*contWhenUniSex*: Continue when there is only one sex in the population. Default to `True`.

**clone()**  
Deep copy of a self mating scheme

### 1.3.8 Class monogamousMating

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=[], weight=0*)  
**REPLENISH** This parameter allows replenishment of one or both parental sex groups in case that they are exhausted. Default to False. Please refer to class `mating` for descriptions of other parameters.

**clone()**  
Deep copy of a random mating scheme

### 1.3.9 Class polygamousMating

A mating scheme of polygyny 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=[], weight=0*)

**FIXME:** No document

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

*polyNum*: Number of sex partners.

*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.

**clone()**  
Deep copy of a random mating scheme

### 1.3.10 Class consanguineousMating

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=[], weight=0*)

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.

*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.

*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`.

**clone()**

Deep copy of a consanguineous mating scheme

### 1.3.11 Class `alphaMating`

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=[], weight=0*)

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 individuals are chosen selectively, and selected again during mating.

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

*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`).

*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`.

**clone()**

Deep copy of a random mating scheme

### 1.3.12 Class `haplodiploidMating`

Haplodiploid mating scheme of many hymenopterans 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=[], weight=0*)

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

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

*alphaNum*: Number of alpha individual. Default to one.

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

**clone()**

Deep copy of a random mating scheme

### 1.3.13 Class `pyMating`

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=[], weight=0*)

Create a Python mating scheme

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

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

**clone()**

Deep copy of a Python mating scheme

### 1.3.14 Class `heteroMating`

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=[], weight=0*)

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

*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 specific virtual subpopulations.

**clone()**

Deep copy of a Python mating scheme

### 1.3.15 Class `sequentialParentChooser`

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

### 1.3.16 Class `sequentialParentsChooser`

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
```

### 1.3.17 Class randomParentChooser

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`, parameter `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)
    FIXME: No document
    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.
    clone ()
        FIXME: No document
```

### 1.3.18 Class randomParentsChooser

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, alpha-
                             Sex=Male, alphaNum=0, alphaField=string)
    Note: If selection is enabled, it works regularly on on-alpha sex, but works twice on alpha sex. That is to say,
    alphaNum alpha individuals are chosen selectively, and selected again during mating.
    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.
    polySex: Male (polygyny) or Female (polyandry) parent that will have polyNum sex partners.
    polyNum: Number of sex partners.
    alphaSex: The sex of the alpha individual, i.e. alpha male or alpha female who be the only mating individuals
    in their sex group.
    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).
    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.
    clone ()
        FIXME: No document
```



### 1.3.19 Class `infoParentsChooser`

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)
```

FIXME: No document

*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.

```
clone ()
```

FIXME: No document

### 1.3.20 Class `pyParentsChooser`

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)
```

FIXME: No document

*parentsGenerator*: A Python generator function

```
clone ()
```

FIXME: No document

```
finalize (pop, sp)
```

FIXME: No document

### 1.3.21 Class `cloneOffspringGenerator`

Clone offspring generator copies parental genotype 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)
```

FIXME: No document

*sexParam*: Ignored because sex is copied from the parent.

*sexMode*: Ignored because sex is copied from the parent.

```
clone ()
```

FIXME: No document

### 1.3.22 Class selfingOffspringGenerator

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 diploid parent. A random copy of the parental chromosomes is chosen randomly to form the paternal 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

### 1.3.23 Class haplodiploidOffspringGenerator

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 considered 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

### 1.3.24 Class mendelianOffspringGenerator

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 controlled 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

# Operator References

## 2.1 Common interface of all operators

### 2.1.1 Class `baseOperator`

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. In the latter case, operators are passed to the `evolve` function of a simulator, and are applied repeatedly during the evolution of the simulator.

The *baseOperator* class is the base class for all operators. It defines a common user interface that specifies at which generations, at which stage of a life cycle, to which populations and subpopulation an operator will be applied. These are achieved by a common set of parameters such as `begin`, `end`, `step`, `at`, `stage` for all operators. Note that a specific operator does not have to honor all these parameters. For example, a recombinator can only be applied during mating so it ignores the `stage` parameter.

An operator can be applied to all or part of the generations during the evolution of a simulator. At the beginning of an evolution, a simulator is usually at the beginning of generation 0. If it evolves 10 generations, it evolves generations 0, 1, ..., and 9 (10 generations) and stops at the beginning of generation 10. A negative generation number `a` has generation number `10 + a`, with `-1` referring to the last evolved generation 9. Note that the starting generation number of a simulator can be changed by its `setGen()` member function.

Output from an operator is usually directed to the standard output (`sys.stdout`). This can be configured using a output specification string, which can be `"` for no output, `'>'` standard terminal output (default), or a filename prefixed by one or more `'>'` characters. In the case of `'>filename'` (or equivalently `'filename'`), the output from an operator is written to this file. However, if two operators write to the same file `filename`, or if an operator write to this file more than once, only the last write operation will succeed. In the case of `'>>filename'`, file `filename` will be opened at the beginning of the evolution and closed at the end. Outputs from multiple operators are appended. `>>>filename` works similar to `>>filename` but `filename`, if it already exists at the beginning of an evolutionary process, will not be cleared.

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

The following parameters can be specified by all operators. However, an operator can ignore some parameters and the exact meaning of a parameter can vary.

*output*: A string that specifies how output from an operator is written, which can be `"` (no output), `'>'` (standard output), or `'filename'` prefixed by one or more `'>'`.

*outputExpr*: An expression that determines the output parameter 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 `>>out10_1.xml` for replicate 1 at generation 10.

*stage*: Stage(s) of a life cycle at which an operator will be applied. It can be `PreMating`, `DuringMating`, `PostMating` and any of their combined stages `PrePostMating`,

`PreDuringMatingDuringPostMating` and `PreDuringPostMating`. Note that all operators have their default stage parameter and some of them ignores this parameter because they can only be applied at certain stage(s) of a life cycle.

*begin*: The starting generation at which an operator will be applied. Default to 0. A negative number is interpreted as a generation counted from the end of an evolution (-1 being the last evolved generation).

*end*: The last generation at which an operator will be applied. Default to -1, namely the last generation.

*step*: The number of generations between applicable generations. Default to 1.

*at*: A list of applicable generations. Parameters *begin*, *end*, and *step* will be ignored if this parameter is specified.

*rep*: A list of applicable replicates. An empty list (default) is interpreted as all replicates in a simulator. Negative indexes such as -1 (last replicate) is acceptable. `rep=idx` can be used as a shortcut for `rep=[idx]`.

*subPop*: A list of applicable (virtual) subpopulations, such as `subPop=[sp1, sp2, (sp2, vsp1)]`. An empty list (default) is interpreted as all subpopulations. `subPop=[sp1]` can be simplified as `subPop=sp1`. Negative indexes are not supported. Support for this parameter vary from operator to operator. Some operators do not support virtual subpopulations and some operators do not support this parameter at all. Please refer to the reference manual of individual operators for their support for this parameter.

*infoFields*: A list of information fields that will be used by an operator. You usually do not need to specify this parameter because operators that use information fields usually have default values for this parameter.

**apply** (*pop*)

Apply an operator to population *pop* directly, without checking its applicability.

**clone** ()

Return a cloned copy of an operator. This function is available to all operators.

## 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*=[], *infoFields*=[])

Create an initializer. Default to be always active.

*subPop*: An array specifies applicable subpopulations

*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 initialize 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*

*atPloidy*: Initialize which copy of chromosomes. Default to all.

**clone** ()

Deep copy of an initializer

### 2.2.2 Class `initSex` (Function `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=[], infoFields=[])
    Initialize individual sex.
    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 turn. If specified, parameter maleFreq is ignored.
    apply (pop)
        Apply this operator to population pop
    clone ()
        Deep copy of an initSex
```

### 2.2.3 Class `initByFreq` (Function `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=[], infoFields=[])
    Randomly assign alleles according to given allele frequencies
    alleleFreq: An array of allele frequencies. The sum of all frequencies must be 1; or for a matrix of allele frequencies, each row corresponds 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.
    apply (pop)
        Apply this operator to population pop
    clone ()
        Deep copy of the operator initByFreq
```

### 2.2.4 Class `initByValue` (Function `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=[]`, `infoFields=[]`)

Initialize a population by given alleles

*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.

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

*maleFreq*: Male frequency

*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`.

**apply** (`pop`)

Apply this operator to population `pop`

**clone** ()

Deep copy of the operator `initByValue`

## 2.2.5 Class `spread` (Function `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=[]`, `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 `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=[]`, `infoFields=[]`)

Initialize populations using given user function

*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`.

*atPloidy*: Initialize which copy of chromosomes. Default to all.

*stage*: Default to `PreMating`.

**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`. Migrants 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*=*subPopList*, *toSubPop*=[], *stage*=`PreMating`, *begin*=0, *end*=-1, *step*=1, *at*=[], *rep*=[], *subPop*=*subPopList*, *infoFields*=["`migrate_to`"])

Create a `migrator`

*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*

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

*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 interpreted as [*sp*]. Note that `fromSubPop=(0, 1)` (two subpopulation) is different from `fromSubPop=[(0,1)]` (a virtual subpopulation).

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

*stage*: Default to `PreMating`

**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=subPopList, toSubPop=[], loci=[], param=None, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=[], subPop=subPopList, infoFields=[]*)

Create a hybrid migrator

*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.

*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.

*stage*: Default to `PreMating`

**apply** (*pop*)

Apply a `pyMigrator`

**clone** ()

Deep copy of a `pyMigrator`



### 2.3.3 Class `splitSubPop` (Function `SplitSubPop`)

Split a subpopulation

**class `splitSubPop`** (*which=0, sizes=[], proportions=[], randomize=True, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=[], subPop=subPopList, infoFields=[]*)

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.

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

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

*proportions*: Proportions of new subpopulations. Should be added up to 1.

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

**`apply`** (*pop*)

Apply a `splitSubPop` operator

**`clone`** ()

Deep copy of a `splitSubPop` operator

### 2.3.4 Class `mergeSubPops` (Function `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=[], subPop=subPopList, infoFields=[]*)

Merge subpopulations

*subPops*: Subpopulations to be merged. Default to all.

**`apply`** (*pop*)

Apply a `mergeSubPops` operator

**`clone`** ()

Deep copy of a `mergeSubPops` operator

### 2.3.5 Class `resizeSubPops` (Function `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 is 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=[], subPop=subPopList, infoFields=[]*)

Resize subpopulations

*newSizes*: Of the specified (or all) subpopulations.

*subPops*: Subpopulations to be resized. Default to all.

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

**`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=[]*, *subPop=subPopList*, *infoFields=[]*)

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.

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

*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()`.

**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 `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=[], subPop=subPopList, infoFields=[])
```

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

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

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

**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 `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=[], subPop=subPopList, infoFields=[])
```

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.

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

**clone()**

Deep copy of a `smmMutator`

### 2.4.4 Class `gsmMutator` (Function `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=[], subPop=subPopList, infoFields=[])
```

Create a `gsmMutator` The 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.

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

*func*: A function that returns the number of steps. This function does not accept any parameter.

**clone()**

Deep copy of a `gsmMutator`

**mutate** (*allele*)

Mutate according to the GSM model

## 2.4.5 Class `pyMutator` (Function `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=[]*, *subPop=subPopList*, *infoFields=[]*)

Create a `pyMutator`

**clone()**

Deep copy of a `pyMutator`

**mutate** (*allele*)

Mutate according to the mixed model

## 2.4.6 Class `pointMutator` (Function `PointMutate`)

Point mutator Mutate specified individuals at specified loci to a specified 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=[]*, *subPop=subPopList*, *infoFields=[]*)

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

*inds*: Individuals who will mutate

*toAllele*: Allele that will be mutate to

**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`

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=[], subPop=subPopList, infoFields=[])
```

Recombine chromosomes from parents

*intensity*: Intensity of recombination. The actual recombination rate between two loci is determined by  $\text{intensity} \times \text{locus distance}$  (between them).

*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.

*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.

*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.

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

*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/repared successfully, or a conversion event if the junction is not resolved/repared. 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 to 15/16 of this parameter. When *convProb* is 1, all recombination events will be conversion events.

*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  $\approx$  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.

*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.

**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)

**produceOffspring** (*parent, off*)

Recombine parental chromosomes of *parent* and pass them to offspring *off*. The homologous chromosomes of *parent* will be recombined twice and form both homologous sets of the offspring, as if *parent* mates with itself (a selfing inheritance model). If sex chromosomes are present, offspring sex will be determined by which sex chromosomes are inherited by *off*. Random sex is assigned to *off* otherwise.

**produceOffspring** (*mom, dad, off*)

Recombine parental chromosomes and pass them to offspring *off*. A Mendelian inheritance model will be used, which recombine homologous sets of chromosomes of *mom* and *dad* and pass them as the first and second sets of homologous chromosomes to offspring *off*, respectively. If sex chromosomes are present, offspring sex is determined by which sex chromosomes are inherited by *off*. Random sex is assigned to *off* otherwise.

**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  $f_i, i = 1, \dots, N$ , 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*=[], *subPop*=*subPopList*, *infoFields*=["fitness"])  
 Create a selector  
*subPop*: A shortcut to `subPops=[subPop]`  
*subPops*: Subpopulations that the selector will apply to. Default to all.  
**apply** (*pop*)  
 Set fitness to all individuals. No selection will happen!  
**clone** ()  
 Deep copy of a selector

## 2.6.2 Class mapSelector (Function MapSelector)

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*=[], *subPop*=*subPopList*, *infoFields*=["fitness"])  
 Create a map selector  
*locus*: The locus index. A shortcut to `loci=[locus]`  
*loci*: The locus indexes. The genotypes at these loci will be used to determine the fitness value.  
*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.  
*phase*: If `True`, genotypes a-b and b-a will have different fitness values. Default to `False`.  
*output*: And other parameters please refer to help (`baseOperator.__init__`)

**clone()**  
 Deep copy of a map selector

**indFitness** (*ind, gen*)  
 Calculate/return the fitness value, currently assuming diploid

### 2.6.3 Class `maSelector` (Function `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=[], subPop=subPopList, infoFields=["fitness"]*)

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

*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.

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

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

#### 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 `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=[], subPop=subPopList, infoFields=["fitness"]*)

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

*selectors*: A list of selectors



```

clone()
    Deep copy of a mlSelector
indFitness (ind, gen)
    Calculate/return the fitness value, currently assuming diploid

```

## 2.6.5 Class pySelector (Function 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=[], sub-
                   Pop=subPopList, infoFields=["fitness"])
    Create a Python hybrid selector

```

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

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

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

*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.

```

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=[], subPop=subPopList, infoFields=[]*)

Create a penetrance operator

*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.

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

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

**`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 `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=[], subPop=subPopList, infoFields=[]*)

Create a map penetrance operator

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

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

*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`.

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

**`clone`** ()

Deep copy of a map penetrance operator

## 2.7.3 Class `maPenetrance` (Function `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=[], subPop=subPopList, infoFields=[]*)  
 Create a multiple allele penetrance operator (penetrance according to diseased or wildtype alleles)  
*locus*: The locus index. The genotype of this locus will be used to determine penetrance.  
*loci*: The locus indexes. The genotypes of these loci will be examined.  
*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.  
*output*: And other parameters please refer to help(baseOperator.\_\_init\_\_)  
**clone** ()  
 Deep copy of a multi-allele penetrance operator  
**penet** (*ind*)  
 Currently assuming diploid

## 2.7.4 Class mlPenetrance (Function MlPenetrance)

Penetrance according to the genotype according to a multiple loci multiplicative model This is the 'multiple-locus' penetrance 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=[], subPop=subPopList, infoFields=[]*)  
 Create a multiple locus penetrance operator  
*peneOps*: A list of penetrance operators  
*mode*: Can be one of PEN\_Multiplicative, PEN\_Additive, and PEN\_Heterogeneity  
**clone** ()  
 Deep copy of a multi-loci penetrance operator  
**penet** (*ind*)  
 Currently assuming diploid

## 2.7.5 Class pyPenetrance (Function 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 genotypes 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=[], subPop=subPopList, infoFields=[]*)  
 Provide locus and penetrance for 11, 12, 13 (in the form of dictionary)

*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.

*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.

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

**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=[], subPop=subPopList, 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 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  $N(0, \sigma^2)$ . 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=[], subPop=subPopList, infoFields=["qtrait"]*)  
 Create a map quantitative trait operator

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

*loci*: An array of locus indexes. The quantitative trait is determined by genotypes at these loci.

*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  $N(\text{mean}, \sigma^2)$ . For multiple loci, the form is 'a-b1-c1-d1-e1-f' etc.

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

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

*output*: And other parameters please refer to help (baseOperator.\_\_init\_\_)

**clone()**

Deep copy of a map quantitative trait operator

**qtrait(ind)**

Currently assuming diploid

### 2.8.3 Class maQuanTrait (Function 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  $N(0, \sigma^2)$  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=[], subPop=subPopList, infoFields=["qtrait"]*)

Create a multiple allele quantitative trait operator Please refer to *quanTrait* for other parameter descriptions.

*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].

*output*: And other parameters please refer to help(baseOperator.\_\_init\_\_)

**clone()**

Deep copy of a multiple allele quantitative trait

**qtrait(ind)**

Currently assuming diploid

### 2.8.4 Class mlQuanTrait (Function 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  $\sigma_i$  (for  $f_i$ ) and  $\sigma$  (for  $f$ ) will be considered. I.e, the trait value should be

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

for QT\_Additive case. If this is not desired, you can set some of the  $\sigma$  to zero.

**class mlQuanTrait** (*qtraits, mode=QT\_Multiplicative, sigma=0, ancestralGen=-1, stage=PostMating, begin=0, end=-1, step=1, at=[], rep=[], subPop=subPopList, infoFields=["qtrait"]*)  
 Create a multiple locus quantitative trait operator Please refer to `quanTrait` for other parameter descriptions.  
*qtraits*: A list of quantitative traits  
*mode*: Can be one of `QT_Multiplicative` and `QT_Additive`  
**clone** ()  
 Deep copy of a multiple loci quantitative trait operator  
**qtrait** (*ind*)  
 Currently assuming diploid

## 2.8.5 Class `pyQuanTrait` (Function `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=[], subPop=subPopList, infoFields=["qtrait"]*)  
 Create a Python quantitative trait operator Please refer to `quanTrait` for other parameter descriptions.  
*loci*: The genotypes at these loci will be passed to *func*.  
*func*: A Python function that accepts genotypes at specified loci and returns the quantitative trait value.  
*output*: And other parameters please refer to `help(baseOperator.__init__)`  
**clone** ()  
 Deep copy of a Python quantitative trait operator  
**qtrait** (*ind*)  
 Currently assuming diploid

## 2.9 Ascertainment

## 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=[], subPop=subPopList, infoFields=[]*)  
 Create a stator  
**clone** ()  
 Deep copy of a stator

### 2.10.2 Class `stat` (Function `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 disequilibrium 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=[], subPop=subPopList, infoFields=[])
```

Create an `stat` operator

**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.

**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 chosen from the following options: `numOfMale`, `propOfMale`, `numOfFemale`, and `propOfFemale`.

**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 chosen from the following options: `numOfAffected`, `propOfAffected`, `numOfUnaffected`, and `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 (array 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`.

*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 (array 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`.

*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`.

*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 calculated 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 `x` 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]`.

*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 chosen from the following options: `subpop` and `midValues`.

*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 sequences on loci `haplo`.

- `haploFreq[haplo], subPop[sp] ['haploFreq'] [haplo]`, the frequency of allele sequences on loci `haplo`.

*LD*: Calculate linkage disequilibria  $LD$ ,  $LD'$  and  $r^2$ , given  $LD=[ [loc1, loc2], [ loc1, loc2, allele1, allele2], \dots ]$ . 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,

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

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

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

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

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'] ['allele1-allele2']`

- `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.

*association*: Association measures

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

*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);
- $\theta$  (  $F_{ST}$  ) 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`.

*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.

*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

*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`.

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

*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.

**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=[], infoFields=[])
```

Dump a population

*alleleOnly*: Only display allele

*infoOnly*: Only display genotypic information

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

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

*chrom*: Chromosome(s) to display

*loci*: Loci to display

*subPop*: Only display subpopulation(s)  
*indRange*: Range(s) of individuals 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\_\_)

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

**apply** (*pop*)  
 Apply an operator to population *pop* directly, without checking its applicability.

**clone** ()  
 Return a cloned copy of an operator. This function is available to all operators.

**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=[]*, *subPop=subPopList*, *infoFields=[]*)

Save population

*output*: Output filename.

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

*format*: Obsolete parameter

*compress*: Obsolete parameter

**apply** (*pop*)  
 Apply an operator to population *pop* directly, without checking its applicability.

**clone** ()  
 Return a cloned copy of an operator. This function is available to all operators.

### 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=[]*, *subPop=subPopList*, *infoFields=[]*)

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

*str*: String to be outputted

**apply** (*pop*)  
 Simply output some info

**clone()**

Return a cloned copy of an operator. This function is available to all operators.

**setString(str)**

Set output string.

#### 2.11.4 Class `pyEval` (Function `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=[]*, *subPop=subPopList*, *infoFields=[]*)

Evaluate expressions/statments in the local namespace of a replicate

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

*stmts*: The statement that will be executed before the expression

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

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

*exposePop*: If `True`, expose the current population as a variable named `pop`

*name*: Used to let pure Python operator to identify themselves

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

**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 `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=[]*, *subPop=subPopList*, *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 `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=[], subPop=subPopList, infoFields=[])
```

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 available 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.

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

*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`

*exposePop*: If `True`, expose the current population as a variable named `pop`

*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.

**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 `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=[], subPop=subPopList, 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 `tager`

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 tager (output="", outputExpr="", stage=DuringMating, begin=0, end=-1, step=1, at=[], rep=[], subPop=subPopList, infoFields=[])  
    Create a tager, default to be always active but no output  
apply (pop)  
    Add a newline  
clone ()  
    Deep copy of a  
    tager
```

### 2.12.2 Class `inheritTagger`

Inherit 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=[], subPop=subPopList, output="", outputExpr="", infoFields=["paternal_tag", "maternal_tag"])  
    Create an inheritTagger that inherits a tag from one or both parents  
mode: Can be one of TAG_Paternal, TAG_Maternal, and TAG_Both  
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=[], subPop=subPopList, 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=[], subPop=subPopList, 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=[], output=">", outputExpr="", stage=PostMating,  
                rep=[], subPop=subPopList, infoFields=[])
    FIXME: No document
    code: Code for Male and Female, default to 1 and 2, respectively. This is used by Linkage format.
    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=[], subPop=subPopList,  

                     stage=PostMating, output=">", outputExpr="", infoFields=[])
```

FIXME: No document

*code*: Code for Male and Female, default to 1 and 2, respectively. This is used by Linkage format.

```
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=[], subPop=subPopList, 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=[], subPop=subPopList, output="", output-  

                 Expr="", infoFields=[])
```

Creates a `pyTagger` that works on specified information fields

*infoFields*: Information fields. The user should guarantee the existence of these fields.

*func*: A Python 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.

```
applyDuringMating (pop, offspring, dad=None, mom=None)
```

Apply the `pyTagger`

```
clone ()
```

Deep copy of a `pyTagger`

## 2.13 Terminator

### 2.13.1 Class `terminateIf`

This operator evaluates an expression in a population's local namespace and terminate the evolution of this population, or the whole simulator, if the return value of this expression is `True`. Termination caused by an operator will stop the execution of all operators after it. Because a life-cycle is considered to be complete if mating is complete, the *evolved generations* (return value from `simulator::evolve`) of a terminated replicate is determined by when the last evolution cycle is terminated.

```
class terminateIf (condition="", stopAll=False, message="", output="", outputExpr="", stage=PostMating, be-  

                  gin=0, end=-1, step=1, at=[], rep=[], subPop=subPopList, infoFields=[])
```

Create a terminator with an expression *condition*, which will be evaluated in a population's local namespace



when the operator is applied to this population. If the return value of *condition* is `True`, the evolution of the population will be terminated. If *stopAll* is set to `True`, the evolution of all replicates of the simulator will be terminated. If this operator is allowed to write to an *output* or *outputExpr* (both default to `""`), the generation number, preceded with an optional *message* will be written to it.

**apply** (*pop*)

Apply an operator to population *pop* directly, without checking its applicability.

**clone** ()

Deep copy of a `terminateIf` 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=[], subPop=subPopList, infoFields=[]*)

Python operator, using a function that accepts a population object.

*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.

*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.

*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`.

#### 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

## 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=[], subPop=subPopList, infoFields=[]*)

A Pre- or PostMating Python operator that apply a function to each individual

*func*: A Python function that accepts an individual and optional genotype and parameters.

*param*: Any Python object that will be passed to `func` after `pop` parameter. Multiple parameters can be passed as a tuple.

*infoFields*: If given, `func` is expected to return an array of the same length and fill these `infoFields` of an individual.

**apply** (*pop*)

Apply the `pyIndOperator` operator to one population

## 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 mechanism 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=[], subPop=subPopList, infoFields=[])
    Create a conditional operator
    cond: Expression that will be treated as a boolean variable
    ifOp: An operator that will be applied when cond is True
    elseOp: An operator that will be applied when cond is False
    apply (pop)
        Apply the ifElse operator to one population
```

### 2.15.2 Class turnOnDebug (Function TurnOnDebug)

Set debug on Turn on debug. There are several ways to turn on debug information for non-optimized modules, namely

- set environment variable SIMUDEBUB.
- 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=[], subPop=subPopList, infoFields=[])
    Create a turnOnDebug operator
```

### 2.15.3 Class turnOffDebug (Function TurnOffDebug)

Set debug off Turn off debug.

```
class turnOffDebug (code, stage=PreMating, begin=0, end=-1, step=1, at=[], rep=[], subPop=subPopList, infoFields=[])
    Create a turnOffDebug operator
    apply (pop)
        Apply the turnOffDebug operator to one population
```

### 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=[], subPop=subPopList, infoFields=[])
    Create a none operator
    apply (pop)
        Apply the noneOp operator to one population
```

### 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=">", output-  
Expr="", stage=PostMating, begin=0, end=-1, step=1, at=[], rep=-1, subPop=subPopList, in-  
foFields=[])
```

Stop a simulation. Press 'q' to exit or any other key to continue.

*prompt*: If `True` (default), print prompt message.

*stopOnKeyStroke*: If `True`, stop only when a key was pressed.

*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`.

**apply** (*pop*)

Apply the pause operator to one population

### 2.15.6 Class `ticToc` (Function `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=[], sub-  
Pop=subPopList, infoFields=[])
```

Create a timer

### 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=[], subPop=subPopList, infoFields=[])
```

Create a `setAncestralDepth` operator

**apply** (*pop*)

Apply the `setAncestralDepth` operator to one population

# 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

### **ListRNG()**

List the names of all available random number generators

### **ListDebugCode()**

List all debug codes

### **LoadPopulation(file)**

Load a population from a file.

### **LoadSimulator(file, matingScheme)**

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 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.

### **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, `RNGmt19937` 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 Module `simuOpt`

### 3.3.1 `simuOpt` functions

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 configuration file, users will be asked to input them interactively.

Because these options are reserved, you can not use them in your `simuPOP` script.

**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 useDefault 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, suppress banner information when simuPOP is loaded.

*debug*: a list of debug code (as 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

## 3.4 Module `simuUtil`

### 3.4.1 Class `randomSample`

This operator draws random individuals from a population repeatedly and forms a number of random samples. These samples can be put in the population's local namespace, or save to disk files. The function form of this operator returns a list of samples directly.

**class** `randomSample` (*size*, *\*args*, *\*\*kwargs*)  
 Draw *size* random samples from a population *times* times. *size* can be a number or a list of numbers. In the former case, individuals are drawn from the whole population and the samples has only one subpopulation. In the latter case, a given number of individuals are drawn from each subpopulation and the result sample has the same number of subpopulation as the population from which samples are drawn. The samples are saved in the population's local namespace if *name* or *nameExpr* is given, and are saved as diskfiles if *saveAs* or *saveAsExpr* is given.

### 3.4.2 Class `caseControlSample`

This operator chooses random cases and controls from a population repeatedly. These samples can be put in the population's local namespace, or save to disk files. The function form of this operator returns a list of samples directly.

**class** `caseControlSample` (*cases*, *controls*, *\*args*, *\*\*kwargs*)  
 Draw *cases* affected and *controls* unaffected individuals from a population repeatedly. *cases* can be a number or a list of numbers. In the former case, affected individuals are drawn from the whole population. In the latter case, a given number of individuals are drawn from each subpopulation. The same hold for *controls*. The resulting samples have two subpopulations that hold cases and controls respectively. The samples are saved in the population's local namespace if *name* or *nameExpr* is given, and are saved as diskfiles if *saveAs* or *saveAsExpr* is given.

### 3.4.3 Class `affectedSibpairSample`

This operator chooses affected sibpairs and their parents from a population repeatedly. These samples can be put in the population's local namespace, or save to disk files. The function form of this operator returns a list of samples directly.

The population to be sampled needs to have at least one ancestral generation. In addition, parents of each offspring is needed so information fields, most likely *father\_idx* and *mother\_idx* should be used to track parents in the parental generation. An during mating operator *parentsTagger* is designed for such a purpose. In addition, because it is very unlikely for two random offspring to share parents, affected sibpairs can only be ascertained from populations that are generated using a mating scheme that produces more than one offspring at each mating event.

**class `affectedSibpairSample`** (*size*, *infoFields*=['father\_idx', 'mother\_idx'], \*args, \*\*kwargs)

Draw *size* families, including two affected siblings and their parents from a population repeatedly. The population to be sampled must have at least one ancestral generation. It should also have two information fields specified by parameter *infoFields* (Default to ['father\_idx', 'mother\_idx']). Parameter *size* can be a number or a list of numbers. In the former case, affected sibpairs are drawn from the whole population. In the latter case, a given number of affected sibpairs are drawn from each subpopulation. In both cases, affected sibpairs in the resulting sample form their own subpopulations (of size two). The samples are saved in the population's local namespace if *name* or *nameExpr* is given, and are saved as diskfiles if *saveAs* or *saveAsExpr* is given.

### 3.4.4 Class `pySubset`

This operator rearranges and removes individuals according to their values at an information field. Individuals with positive values at this information field are moved to the subpopulation specified by the integer value of this value. Individuals with negative values are removed. There is no function form of this operator because this operator is essentially a wrapper around function `population::setSubPopByIndInfo(field)`.

**class `pySubset`** (*field*, \*args, \*\*kwargs)

Create a `pySubset` operator that rearranges and removes individuals according to their values at an information field *field*.

### 3.4.5 Class `simuProgress`

This class defines a very simple text based progress bar. It will display a character (default to ".") for each change of progress (default to 2%), and a number (1, 2, ..., 9) for each 10% of progress, and print a message (default to "Done.\n") when the job is finished.

This class is used as follows:

```
progress = simuProgress("Start simulation", 500)
for i in range(500):
    progress.update(i+1)
# if you would like to make sure the done message is displayed.
progress.done()
```

**class `simuProgress`** (*message*, *totalCount*, *progressChar*='.', *block*=2, *done*='Done.  
n')

*totalCount*: Total expected steps.

*progressChar*: Character to be displayed for each progress.

*block*: display progress at which interval (in terms of percentage)?

*done*: Message displayed when the job is finished.

```
done ()
    Finish progressbar, print 'done' message.

update (count)
    Update the progreebar.
```

### 3.4.6 simuUtil functions

This module provides some commonly used operators and format conversion utilities.

#### **AffectedSibpairSample** (*pop*, *size*, *\*args*, *\*\*kwargs*)

Function version of operator affectedSibpairSample whose `__init__` function is

Draw *size* families, including two affected siblings and their parents from a population repeatedly. The population to be sampled must have at least one ancestral generation. It should also have two information fields specified by parameter *infoFields* (Default to [`'father_idx'`, `'mother_idx'`]). Parameter *size* can be a number or a list of numbers. In the former case, affected sibpairs are drawn from the whole population. In the latter case, a given number of affected sibpairs are drawn from each subpopulation. In both cases, affected sibpairs in the resulting sample form their own subpopulations (of size two). The samples are saved in the population's local namespace if *name* or *nameExpr* is given, and are saved as diskfiles if *saveAs* or *saveAsExpr* is given.

#### **CaseControlSample** (*pop*, *\*args*, *\*\*kwargs*)

Function version of operator caseControlSample whose `__init__` function is

Draw *cases* affected and *controls* unaffected individuals from a population repeatedly. *cases* can be a number or a list of numbers. In the former case, affected individuals are drawn from the whole population. In the latter case, a given number of individuals are drawn from each subpopulation. The same hold for *controls*. The resulting samples have two subpopulations that hold cases and controls respectively. The samples are saved in the population's local namespace if *name* or *nameExpr* is given, and are saved as diskfiles if *saveAs* or *saveAsExpr* is given.

#### **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 *initSize* to *endSize* after *burnin*, split the population at generation *split*.

**InstantExpansion** (*initSize, endSize, end, burnin=0, split=0, numSubPop=1, bottleneckGen=-1, bottleneckSize=0*)

Instantaneously expand population size from *initSize* 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 *initSize* 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

$$\begin{matrix} & x & m/(n-1) & m/(n-1) & \dots \\ m/(n-1) & x & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & m/(n-1) & m/(n-1) & x & \dots \end{matrix}$$

where  $x = 1-m$

**MigrSteppingStoneRates** (*r, n, circular=False*)

migration rate matrix, circular stepping stone model ( $X=1-m$ )

$$\begin{matrix} X & m/2 & & & m/2 \\ m/2 & X & m/2 & & 0 \\ 0 & m/2 & x & m/2 & \dots \\ \dots & \dots & \dots & \dots & \dots \\ m/2 & 0 & \dots & m/2 & X \end{matrix}$$

or non-circular

$$\begin{matrix} X & m/2 & & & m/2 \\ m/2 & X & m/2 & & 0 \\ 0 & m/2 & X & m/2 & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & & & m & X \end{matrix}$$

**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.

**RandomSample** (*pop*, *\*args*, *\*\*kwargs*)

Function version of operator randomSample whose `__init__` function is

Draw *size* random samples from a population *times* times. *size* can be a number or a list of numbers. In the former case, individuals are drawn from the whole population and the samples has only one subpopulation. In the latter case, a given number of individuals are drawn from each subpopulation and the result sample has the same number of subpopulation as the population from which samples are drawn. The samples are saved in the population's local namespace if *name* or *nameExpr* is given, and are saved as diskfiles if *saveAs* or *saveAsExpr* is given.

**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

*fields*: 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

## 3.5 Module `simuRPy`

### 3.5.1 Class `varPlotter`

This class defines a Python operator that uses R to plot a `simuPOP` express. During the evolution, this express is evaluated in each replicate's local namespace. How this expression is plotted depends on the dimension of the return value (if a sequence is returned), number of replicates, whether or not historical values (collected over several generations) are plotted, and plot type (lines or images).

The default behavior of this operator is to plot the history of an expression. For example, when operator

```
varPlotter(var='expr')
```

is used in `simulator::evolve`, the value of `expr` will be recorded each time when this operator is applied. A line will be draw in a figure with x-axis being the generation number. Parameters `ylim` can be used to specify the range of y-axis.

If the return value of expression `expr` is a sequence (tuple or list), parameter `varDim` has to be used to indicate the dimension of this expression. For example,

```
varPlotter(var='expr', varDim=3)
```

will plot three lines, corresponding to the histories of each item in the array.

If the expression returns a number and there are several replicates, parameter `numRep` should be used. In this case, each line will correspond to a replicate.

If the expression returns a vector and there are several replicates, several subplots will be used. Parameter `byRep` or `byVar` should be used to tell `varPlotter` whether the subplots should be divided by replicate or by variable. For example,

```
varPlotter(var='expr', varDim=8, numRep=5, byRep=1)
```

will use an appropriate layout for your subplots, which is, in this case, 2x3 for 5 replicates. Each subplot will have 8 lines. If `byVal` is `True`, there will be 3x3 subplots for 8 items in an array, and each subplot will have 5 lines. Note that `byRep` or `byVal` can also be used when there is only one replicate or if the dimension of the expression is one.

When `history=False`, histories of each variable will be discarded so the figure will always plot the current value of the expression.

```
class varPlotter(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)
```

*expr*: expression that will be evaluate at each replicate's local namespace when the operator is applied.

*history*: whether or not record and plot the history of an expression. Default to `True`.

*varDim*: If the return value of `expr` is a sequence, `varDim` should be set to the length of this sequence. Default to 1.

*numRep*: Number of replicates of the simulator. Default to 1.

*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 of an expression. The default value is 0, which will keep all histories.

*ylim*: The range of y-axis.

*update*: Update figure after update generations. This is used when you do not want to update the figure every time when this operator is applied.

*title, xlab, ylab*: Title, label at x and y axes of your figure(s). *xtitle* is defaulted to 'generation'.

*axes*: Whether or not plot axes. Default to `True`.

*lty*: A list of line type for each line in the figure.

*col*: A list of colors for each line in the figure.

*level*: level of image colors (default to 20).

*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.

*leaveOpen*: whether or not leave the plot open when plotting is done. Default to `True`.



# INDEX

absIndIndex () (population method), 8  
absLocusIndex () (GenoStruTrait method), 4  
addChrom () (population method), 10  
addChromFromPop () (population method), 10  
addIndFromPop () (population method), 10  
addInfoField () (population method), 11  
addInfoField () (simulator method), 13  
addInfoFields () (population method), 11  
addInfoFields () (simulator method), 13  
addLoci () (population method), 10  
addLociFromPop () (population method), 10  
affected () (individual method), 6  
affectedChar () (individual method), 6  
AffectedSibpairSample () (in module ), 68  
affectedSibpairSample (class in ), 67  
affectionSplitter (class in ), 15  
affectionTagger (class in ), 56  
allele () (individual method), 5  
alleleName () (GenoStruTrait method), 5  
alleleNames () (GenoStruTrait method), 5  
alleleOnly () (dumper method), 51  
AlleleType () (in module ), 61  
alphaMating (class in ), 22  
ancestor () (population method), 8, 9  
ancestralGens () (population method), 9  
apply () (affectionTagger method), 56  
apply () (baseOperator method), 28  
apply () (dumper method), 51  
apply () (ifElse method), 59  
apply () (infoEval method), 53  
apply () (infoTagger method), 56  
apply () (initByFreq method), 29  
apply () (initByValue method), 30  
apply () (initSex method), 29  
apply () (mergeSubPops method), 33  
apply () (migrator method), 32  
apply () (mutator method), 34  
apply () (noneOp method), 59  
apply () (parentTagger method), 54  
apply () (parentsTagger method), 55  
apply () (pause method), 60  
apply () (penetrance method), 42  
apply () (pointMutator method), 36  
apply () (pyEval method), 52  
apply () (pyIndOperator method), 58  
apply () (pyInit method), 31  
apply () (pyMigrator method), 32  
apply () (pyOperator method), 58  
apply () (pyOutput method), 51  
apply () (quanTrait method), 44  
apply () (resizeSubPops method), 33  
apply () (savePopulation method), 51  
apply () (selector method), 39  
apply () (setAncestralDepth method), 60  
apply () (sexTagger method), 55  
apply () (splitSubPop method), 33  
apply () (spread method), 30  
apply () (stat method), 50  
apply () (tagger method), 54  
apply () (terminateIf method), 57  
apply () (turnOffDebug method), 59  
applyDuringMating () (inheritTagger method), 54  
applyDuringMating () (parentTagger method), 55  
applyDuringMating () (parentsTagger method), 55  
applyDuringMating () (pyTagger method), 56  
ascertainment, 46  
  
baseOperator (class in ), 27  
baseRandomMating (class in ), 20  
binomialSelection (class in ), 19  
  
CaseControl\_ChiSq () (in module ), 68  
CaseControlSample () (in module ), 68  
caseControlSample (class in ), 66  
ChiSq\_test () (in module ), 68  
chromBegin () (GenoStruTrait method), 3  
chromByName () (GenoStruTrait method), 4  
chromEnd () (GenoStruTrait method), 4  
chromLocusPair () (GenoStruTrait method), 4  
chromName () (GenoStruTrait method), 4  
chromNames () (GenoStruTrait method), 4  
chromType () (GenoStruTrait method), 4  
chromTypes () (GenoStruTrait method), 4  
class

- affectionSplitter, 15
- affectionTagger, 55
- alphaMating, 22
- baseOperator, 27
- baseRandomMating, 19
- binomialSelection, 19
- cloneMating, 19
- cloneOffspringGenerator, 25
- combinedSplitter, 17
- consanguineousMating, 21
- dumper, 50
- GenoStruTrait, 3
- genotypeSplitter, 17
- gsmMutator, 35
- haplodiploidMating, 22
- haplodiploidOffspringGenerator, 26
- heteroMating, 23
- ifElse, 58
- individual, 5
- infoEval, 52
- infoExec, 53
- infoParentsChooser, 25
- infoSplitter, 16
- infoTagger, 56
- inheritTagger, 54
- initByFreq, 29
- initByValue, 29
- initializer, 28
- initSex, 29
- kamMutator, 35
- maPenetrance, 42
- mapPenetrance, 42
- mapQuanTrait, 44
- mapSelector, 39
- maQuanTrait, 45
- maSelector, 40
- mating, 17
- mendelianOffspringGenerator, 26
- mergeSubPops, 33
- migrator, 31
- mlPenetrance, 43
- mlQuanTrait, 45
- mlSelector, 40
- monogamousMating, 21
- mutator, 34
- noMating, 18
- noneOp, 59
- parentsTagger, 55
- parentTagger, 54
- pause, 60
- pedigree, 14
- penetrance, 41
- pointMutator, 36
- polygamousMating, 21

- population, 7
- proportionSplitter, 16
- pyEval, 52
- pyExec, 52
- pyIndOperator, 58
- pyInit, 30
- pyMating, 23
- pyMigrator, 32
- pyMutator, 36
- pyOperator, 57
- pyOutput, 51
- pyParentsChooser, 25
- pyPenetrance, 43
- pyQuanTrait, 46
- pySelector, 41
- pyTagger, 56
- quanTrait, 44
- randomMating, 20
- randomParentChooser, 24
- randomParentsChooser, 24
- rangeSplitter, 16
- recombinator, 37
- resizeSubPops, 33
- RNG, 62
- savePopulation, 51
- selector, 38
- selfingOffspringGenerator, 26
- selfMating, 20
- sequentialParentChooser, 23
- sequentialParentsChooser, 23
- setAncestralDepth, 60
- sexSplitter, 15
- sexTagger, 55
- simulator, 12
- smmMutator, 35
- splitSubPop, 33
- spread, 30
- stat, 46
- stator, 46
- tagger, 54
- terminateIf, 56
- ticToc, 60
- turnOffDebug, 59
- turnOnDebug, 59
- vspSplitter, 15
- clone() (alphaMating method), 22
- clone() (baseOperator method), 28
- clone() (baseRandomMating method), 20
- clone() (binomialSelection method), 19
- clone() (cloneMating method), 19
- clone() (cloneOffspringGenerator method), 25
- clone() (consanguineousMating method), 22
- clone() (dumper method), 51
- clone() (gsmMutator method), 36

clone () (haplodiploidMating method), 23  
 clone () (haplodiploidOffspringGenerator method), 26  
 clone () (heteroMating method), 23  
 clone () (infoEval method), 53  
 clone () (infoExec method), 53  
 clone () (infoParentsChooser method), 25  
 clone () (inheritTagger method), 54  
 clone () (initByFreq method), 29  
 clone () (initByValue method), 30  
 clone () (initSex method), 29  
 clone () (initializer method), 28  
 clone () (kamMutator method), 35  
 clone () (maPenetrance method), 43  
 clone () (maQuanTrait method), 45  
 clone () (maSelector method), 40  
 clone () (mapPenetrance method), 42  
 clone () (mapQuanTrait method), 45  
 clone () (mapSelector method), 40  
 clone () (mating method), 18  
 clone () (mendelianOffspringGenerator method), 26  
 clone () (mergeSubPops method), 33  
 clone () (migrator method), 32  
 clone () (mlPenetrance method), 43  
 clone () (mlQuanTrait method), 46  
 clone () (mlSelector method), 41  
 clone () (monogamousMating method), 21  
 clone () (mutator method), 34  
 clone () (noMating method), 18  
 clone () (parentTagger method), 55  
 clone () (parentsTagger method), 55  
 clone () (penetrance method), 42  
 clone () (pointMutator method), 36  
 clone () (polygamousMating method), 21  
 clone () (population method), 8  
 clone () (pyEval method), 52  
 clone () (pyExec method), 52  
 clone () (pyInit method), 31  
 clone () (pyMating method), 23  
 clone () (pyMigrator method), 32  
 clone () (pyMutator method), 36  
 clone () (pyOutput method), 52  
 clone () (pyParentsChooser method), 25  
 clone () (pyPenetrance method), 44  
 clone () (pyQuanTrait method), 46  
 clone () (pySelector method), 41  
 clone () (pyTagger method), 56  
 clone () (quanTrait method), 44  
 clone () (randomMating method), 20  
 clone () (randomParentChooser method), 24  
 clone () (randomParentsChooser method), 24  
 clone () (recombinator method), 38  
 clone () (resizeSubPops method), 34  
 clone () (savePopulation method), 51  
 clone () (selector method), 39  
 clone () (selfMating method), 21  
 clone () (selfingOffspringGenerator method), 26  
 clone () (sequentialParentChooser method), 23  
 clone () (sequentialParentsChooser method), 24  
 clone () (simulator method), 12  
 clone () (simmMutator method), 35  
 clone () (splitSubPop method), 33  
 clone () (spread method), 30  
 clone () (stat method), 50  
 clone () (stator method), 46  
 clone () (tagger method), 54  
 clone () (terminateIf method), 57  
 clone () (vspSplitter method), 15  
 cloneMating (class in ), 19  
 cloneOffspringGenerator (class in ), 25  
 combinedSplitter (class in ), 17  
 consanguineousMating (class in ), 21  
 ConstSize () (in module ), 68  
 convCount () (recombinator method), 38  
 convCounts () (recombinator method), 38  
 conversion, 37  
 copyParentalGenotype () (haplodiploidOffspringGenerator method), 26  
 done () (simuProgress method), 68  
 dumper (class in ), 50  
 dvars () (population method), 12  
 dvars () (simulator method), 13, 14  
 evolve () (simulator method), 12  
 ExponentialExpansion () (in module ), 69  
 extract () (population method), 10  
 extract () (simulator method), 13  
 finalize () (pyParentsChooser method), 25  
 formOffspringGenotype () (mendelianOffspringGenerator method), 26  
 function  
     GsmMutate, 35  
     InitByFreq, 29  
     InitByValue, 29  
     InitSex, 29  
     KamMutate, 35  
     MaPenetrance, 42  
     MaQuanTrait, 45  
     MaSelect, 40  
     MapPenetrance, 42  
     MapQuanTrait, 44  
     MapSelector, 39  
     MergeSubPops, 33  
     MlPenetrance, 43  
     MlQuanTrait, 45  
     MlSelect, 40  
     PointMutate, 36  
     PyEval, 52

- PyExec, 52
- PyInit, 30
- PyMutate, 36
- PyPenetrance, 43
- PyQuanTrait, 46
- PySelect, 41
- ResizeSubPops, 33
- SmmMutate, 35
- SplitSubPop, 33
- Spread, 30
- Stat, 46
- TicToc, 60
- TurnOffDebug, 59
- TurnOnDebug, 59
- infoEval, 52
- infoExec, 53
- gen () (simulator method), 12
- GenoStruTrait (class in ), 3
- genotype () (individual method), 6
- genotype () (population method), 9
- genotypeSplitter (class in ), 17
- getParam () (in module ), 63
- gsmMutator (class in ), 35
- haplodiploidMating (class in ), 22
- haplodiploidOffspringGenerator (class in ), 26
- heteroMating (class in ), 23
- ifElse (class in ), 59
- indFitness () (maSelector method), 40
- indFitness () (mapSelector method), 40
- indFitness () (mlSelector method), 41
- indFitness () (pySelector method), 41
- indInfo () (population method), 11
- individual, 3
- individual () (population method), 9
- individual (class in ), 5
- individuals () (population method), 9
- info () (individual method), 6
- infoEval (class in ), 53
- infoExec (class in ), 53
- infoField () (GenoStruTrait method), 5
- infoFields () (GenoStruTrait method), 5
- infoIdx () (GenoStruTrait method), 5
- infoOnly () (dumper method), 51
- infoParentsChooser (class in ), 25
- infoSplitter (class in ), 16
- infoTagger (class in ), 56
- inheritTagger (class in ), 54
- initByFreq (class in ), 29
- initByValue (class in ), 30
- initializer, 28
- initializer (class in ), 28
- initSex (class in ), 29
- InstantExpansion () (in module ), 69
- intInfo () (individual method), 6
- kamMutator (class in ), 35
- LargePeds\_Reg\_merlin () (in module ), 69
- LargePeds\_VC\_merlin () (in module ), 69
- Limits () (in module ), 61
- LinearExpansion () (in module ), 70
- ListDebugCode () (in module ), 61
- ListRNG () (in module ), 61
- ListVars () (in module ), 70
- LoadFstat () (in module ), 70
- LoadPopulation () (in module ), 61
- LoadSimulator () (in module ), 61
- locateRelatives () (pedigree method), 14
- lociByNames () (GenoStruTrait method), 4
- lociDist () (GenoStruTrait method), 4
- lociNames () (GenoStruTrait method), 4
- lociPos () (GenoStruTrait method), 4
- locusByName () (GenoStruTrait method), 4
- locusName () (GenoStruTrait method), 4
- locusPos () (GenoStruTrait method), 4
- LOD\_gh () (in module ), 69
- LOD\_merlin () (in module ), 69
- maPenetrance (class in ), 43
- mapPenetrance (class in ), 42
- mapQuanTrait (class in ), 44
- mapSelector (class in ), 39
- maQuanTrait (class in ), 45
- maSelector (class in ), 40
- mating (class in ), 18
- mating scheme, 17
- max () (RNG method), 62
- MaxAllele () (in module ), 61
- maxAllele () (mutator method), 34
- maxSeed () (RNG method), 62
- mendelianOffspringGenerator (class in ), 26
- MergePopulations () (in module ), 61
- MergePopulationsByLocs () (in module ), 61
- mergeSubPops () (population method), 10
- mergeSubPops (class in ), 33
- migrator, 31
- migrator (class in ), 31
- MigrIslandRates () (in module ), 70
- MigrSteppingStoneRates () (in module ), 70
- mlPenetrance (class in ), 43
- mlQuanTrait (class in ), 46
- mlSelector (class in ), 40
- module
  - affectedSibpairSample, 67
  - caseControlSample, 66
  - pySubset, 67



- randomSample, 66
- simuOpt, 63
- simuProgress, 67
- simuUtil, 68
- varPlotter, 74
- ModuleCompiler() (in module ), 61
- ModuleDate() (in module ), 61
- ModulePlatform() (in module ), 61
- ModulePyVersion() (in module ), 61
- monogamousMating (class in ), 21
- mutate() (gsmMutator method), 36
- mutate() (kamMutator method), 35
- mutate() (mutator method), 34
- mutate() (pyMutator method), 36
- Mutation, 34
- mutationCount() (mutator method), 34
- mutationCount() (pointMutator method), 36
- mutationCounts() (mutator method), 34
- mutationCounts() (pointMutator method), 36
- mutator (class in ), 34
- name() (RNG method), 62
- name() (affectionSplitter method), 15
- name() (combinedSplitter method), 17
- name() (genotypeSplitter method), 17
- name() (infoEval method), 53
- name() (infoSplitter method), 16
- name() (proportionSplitter method), 16
- name() (pyEval method), 52
- name() (rangeSplitter method), 16
- name() (sexSplitter method), 15
- name() (vspSplitter method), 15
- noMating (class in ), 18
- noneOp (class in ), 59
- numChrom() (GenoStruTrait method), 4
- numLoci() (GenoStruTrait method), 4
- numRep() (simulator method), 13
- numSubPop() (population method), 8
- numVirtualSubPop() (affectionSplitter method), 16
- numVirtualSubPop() (combinedSplitter method), 17
- numVirtualSubPop() (genotypeSplitter method), 17
- numVirtualSubPop() (infoSplitter method), 16
- numVirtualSubPop() (population method), 8
- numVirtualSubPop() (proportionSplitter method), 16
- numVirtualSubPop() (rangeSplitter method), 16
- numVirtualSubPop() (sexSplitter method), 15
- numVirtualSubPop() (vspSplitter method), 15
- Optimized() (in module ), 61
- parentsTagger (class in ), 55
- parentTagger (class in ), 54
- pause (class in ), 60
- pedigree (class in ), 14
- penet() (maPenetrance method), 43
- penet() (mlPenetrance method), 43
- penet() (penetrance method), 42
- penet() (pyPenetrance method), 44
- penetrance, 41
- penetrance (class in ), 42
- ploidy() (GenoStruTrait method), 3
- ploidyName() (GenoStruTrait method), 3
- pointMutator (class in ), 36
- polygamousMating (class in ), 21
- popSize() (population method), 8
- population, 3
- population() (simulator method), 13
- population (class in ), 7
- populations() (simulator method), 13
- prettyOutput() (in module ), 64
- printConfig() (in module ), 64
- produceOffspring() (recombinator method), 38
- proportionSplitter (class in ), 16
- push() (population method), 9
- pvalChiSq() (RNG method), 62
- pyEval (class in ), 52
- pyExec (class in ), 52
- pyIndOperator (class in ), 58
- pyInit (class in ), 30
- pyMating (class in ), 23
- pyMigrator (class in ), 32
- pyMutator (class in ), 36
- pyOperator (class in ), 57
- pyOutput (class in ), 51
- pyParentsChooser (class in ), 25
- pyPenetrance (class in ), 44
- pyQuanTrait (class in ), 46
- pySelector (class in ), 41
- pySubset (class in ), 67
- pyTagger (class in ), 56
- qtrait() (maQuanTrait method), 45
- qtrait() (mapQuanTrait method), 45
- qtrait() (mlQuanTrait method), 46
- qtrait() (pyQuanTrait method), 46
- qtrait() (quanTrait method), 44
- QtraitSibs\_Reg\_merlin() (in module ), 70
- QtraitSibs\_VC\_merlin() (in module ), 71
- quantitative trait, 44
- quanTrait (class in ), 44
- randBinomial() (RNG method), 62
- randExponential() (RNG method), 62
- randGeometric() (RNG method), 62
- randGet() (RNG method), 62
- randInt() (RNG method), 62
- randMultinomial() (RNG method), 62

randMultinomialVal() (RNG method), 62  
 randNormal() (RNG method), 63  
 randomMating (class in ), 20  
 randomParentChooser (class in ), 24  
 randomParentsChooser (class in ), 24  
 RandomSample() (in module ), 71  
 randomSample (class in ), 66  
 randPoisson() (RNG method), 63  
 randUniform01() (RNG method), 63  
 rangeSplitter (class in ), 16  
 rate() (migrator method), 32  
 rate() (mutator method), 34  
 recCount() (recombinator method), 38  
 recCounts() (recombinator method), 38  
 recombination, 37  
 recombinator (class in ), 37  
 Regression\_merlin() (in module ), 71  
 removeEmptySubPops() (population method), 10  
 removeIndividuals() (population method), 10  
 removeLoci() (population method), 10  
 removeSubPops() (population method), 10  
 requireRevision() (in module ), 65  
 resize() (population method), 11  
 resizeSubPops (class in ), 33  
 RNG (class in ), 62  
 rng() (in module ), 62  
  
 save() (population method), 8  
 save() (simulator method), 12  
 saveConfig() (in module ), 65  
 SaveCSV() (in module ), 71  
 SaveFstat() (in module ), 71  
 saveFstat() (in module ), 73  
 SaveLinkage() (in module ), 71  
 saveLinkage() (in module ), 73  
 SaveMerlinDatFile() (in module ), 72  
 SaveMerlinMapFile() (in module ), 72  
 SaveMerlinPedFile() (in module ), 72  
 savePopulation (class in ), 51  
 SaveQTDT() (in module ), 72  
 SaveSolarFrqFile() (in module ), 72  
 seed() (RNG method), 63  
 selection, 38  
 selector (class in ), 39  
 selfingOffspringGenerator (class in ), 26  
 selfMating (class in ), 20  
 sequentialParentChooser (class in ), 23  
 sequentialParentsChooser (class in ), 24  
 setAffected() (individual method), 6  
 setAllele() (individual method), 5  
 setAlleleOnly() (dumper method), 51  
 setAncestralDepth() (population method), 9  
 setAncestralDepth() (simulator method), 13  
 setAncestralDepth (class in ), 60  
  
 setGen() (simulator method), 12  
 setGenotype() (individual method), 6  
 setGenotype() (population method), 9  
 setIndexesOfRelatives() (pedigree method), 14  
 setIndInfo() (population method), 11  
 setInfo() (individual method), 6  
 setInfoFields() (population method), 12  
 setInfoOnly() (dumper method), 51  
 setMatingScheme() (simulator method), 13  
 setMaxAllele() (mutator method), 34  
 setOptions() (in module ), 65  
 setRate() (mutator method), 34  
 setRates() (migrator method), 32  
 SetRNG() (in module ), 61  
 setRNG() (RNG method), 63  
 setSeed() (RNG method), 63  
 setSex() (individual method), 6  
 setString() (pyOutput method), 52  
 setSubPopByIndInfo() (population method), 11  
 setVirtualSplitter() (population method), 9  
 sex() (individual method), 6  
 sexChar() (individual method), 6  
 sexSplitter (class in ), 15  
 sexTagger (class in ), 55  
 Sibpair\_LOD\_gh() (in module ), 72  
 Sibpair\_LOD\_merlin() (in module ), 73  
 Sibpair\_TDT\_gh() (in module ), 73  
 Simulator, 3  
 simulator (class in ), 12  
 simuProgress (class in ), 67  
 simuRev() (in module ), 62  
 simuVer() (in module ), 62  
 smmMutator (class in ), 35  
 splitSubPop() (population method), 11  
 splitSubPop (class in ), 33  
 spread (class in ), 30  
 stat (class in ), 47  
 stator (class in ), 46  
 submitScratch() (mating method), 18  
 subPopBegin() (population method), 8  
 subPopEnd() (population method), 8  
 subPopIndPair() (population method), 8  
 subPopSize() (population method), 8  
 subPopSizes() (population method), 8  
  
 tagger (class in ), 54  
 TDT\_gh() (in module ), 73  
 terminateIf (class in ), 56  
 ticToc (class in ), 60  
 totNumLoci() (GenoStruTrait method), 4  
 TurnOffDebug() (in module ), 62  
 turnOffDebug (class in ), 59  
 TurnOnDebug() (in module ), 62  
 turnOnDebug (class in ), 59

update() (simuProgress method), 68  
usage() (in module ), 65  
useAncestralGen() (population method), 9  
  
valueAnd() (in module ), 65  
valueBetween() (in module ), 65  
valueEqual() (in module ), 65  
valueGE() (in module ), 65  
valueGT() (in module ), 65  
valueIsList() (in module ), 65  
valueIsNum() (in module ), 65  
valueLE() (in module ), 65  
valueListOf() (in module ), 66  
valueLT() (in module ), 65  
valueNot() (in module ), 66  
valueNotEqual() (in module ), 66  
valueOneOf() (in module ), 66  
valueOr() (in module ), 66  
valueTrueFalse() (in module ), 66  
valueValidDir() (in module ), 66  
valueValidFile() (in module ), 66  
varPlotter (class in ), 74  
vars() (population method), 12  
vars() (simulator method), 13  
VC\_merlin() (in module ), 73  
virtualSubPopName() (population method), 9  
vspSplitter (class in ), 15