

Compilation

The program should run on Linux or Mac OS X. It requires the GNU Scientific Library installed (gnu.org/software/gsl/), and the source code files 'main_dfuse.c', 'func_dfuse.c' and 'head_dfuse.h'. The program can be compiled with

```
'gcc -O2 -Wall -o dfuse main_dfuse.c func_dfuse.c -lgsl -lgslcblas -lm'.
```

Running the program

The program runs fully on default options by typing './dfuse'. Optional arguments can be viewed with './dfuse -h'. Settings for the spatial grid, epistatic loci, non-epistatic loci, and environment may be provided as external input files. Several error checks and warnings are included while running the program, but these do not cover all possible errors, e.g. arising from misspecifications of optional arguments or external input files.

Options

- d deme settings input filename
- e epistatic selection input filename
- s single-locus selection input filename
- v environment settings input filename
- o output filename [default = 'testing']
- r number of replicates [default = 1]
- g number of generations [default = 10]
- G print every x generations [default = 10]
- O print stats for selected loci every generation: 0 (no) or 1 (yes) [default = 0]
- m migration rate [default = 0.05]
- c selection coefficient [default = 0.0]
- a selection against ancestral alleles on DMI ridge [default = 0.0]
- R recessivity for DMIs: 0 (dominant), 1 (co-dominant), 2 (partially recessive), 3 (fully recessive) [default = 0]
- S selection stage 'viability', 'fertility', 'both' or 'none' [default = 'viability']
- M migration stage 'pollen', 'progeny' or 'both' [default = 'progeny']
- D dispersal mode: 0 (infinite source) or 1 (finite source) [default = 0]
- l marker loci per chromosome [default = 51]
- f fitness accumulation 'mult' or 'add' [default = 'mult']
- Y generation to start staining [default = 0 (no staining)]
- h display options

Example command

```
dfuse -d demefile_2 -e epifile_1 -r 20 -g 100 -m 0.2 -c 0.9
```

Inputfiles

Deme settings (-d)

If not specified, the default is to have three demes with adult carrying capacity $N_c = 50$ and progeny capacity $N_p = 100$. If an external input file is specified, the first line has to give the number of demes, followed by a line that specifies the adult carrying capacity for each deme, and a line that specifies the progeny carrying capacity for each deme.

Example: demefile_2

```

3
500 500 500
1000 1000 1000

```

Epistatic selection (-e)

If not specified, there are no defaults, and there won't be epistatic selection. If an external input file is specified, the first line has to give the total number of epistatically interacting loci, the number of epistatic complexes, and the number of loci within each complex (only one value, equal for all complexes). This is followed by several lines that specify loci positions and the type of selection. All lines are required but line 8 will be ignored for pathways.

<i>Line 1:</i>	loci, complexes, loci per complex	
<i>Line 2:</i>	locus count	(count starts at 0)
<i>Line 3:</i>	complex count	(count starts at 0)
<i>Line 4:</i>	locus within complex	(count starts at 0)
<i>Line 5:</i>	selection type	(0 for DMI; 1 for pathway)
<i>Line 6:</i>	chromosome	(count starts at 0)
<i>Line 7:</i>	position on chromosome	(float, between 0 and 1)
<i>Line 8:</i>	ancestral allele	(0 for allele <i>a</i> ; 1 for allele <i>b</i>)

Example: epifile_1

```

2 1 2
0 1
0 0
0 1
0 0
0 0
0.4 0.6
0 1

```

Single-locus selection (-s)

If not specified, there are no defaults, and there won't be single-locus (non-epistatic) selection. If an external input file is specified, the first line has to give the total number of

single-selected loci. This is followed by several lines that specify loci positions and genotypic effects. All lines are required but lines 6 – 8 are only relevant for environmental selection. Environmental selection additionally requires that demes have their environments and environment-specific selection coefficients specified. These are set automatically using default values, or may be provided using an external file (below: Environmental settings).

<i>Line 1:</i>	loci	
<i>Line 2:</i>	locus count	(count starts at 0)
<i>Line 3:</i>	selection type	(1 environmental; 2 overdominance)
<i>Line 4:</i>	chromosome	(count starts at 0)
<i>Line 5:</i>	position on chromosome	(float, between 0 and 1)
<i>Line 6:</i>	<i>aa</i> genotypic effect	(float, between 0 and 1)
<i>Line 7:</i>	<i>ab</i> genotypic effect	(float, between 0 and 1)
<i>Line 8:</i>	<i>bb</i> genotypic effect	(float, between 0 and 1)

Example: `selfile_2`

```

2
0 1
1 1
0 0
0.4 0.6
0.0 1.0
0.5 0.5
1.0 0.0

```

Environmental settings (-v)

If no external file is provided, default values will be used if environmental selection is specified (above: Single-locus selection). As default, the environment will change gradually from 0 to 1 along demes. The selection coefficient can be set with the `-c` flag. If an external input file is specified, the first line has to give the number of environmentally selected loci and the number of demes. For each selected locus, two additional lines specify the settings for each deme: the first one gives the environmental value (between 0.0 and 1.0), and the second one the environmental selection coefficient (these are mandatory and are used instead of the `-c` specified value for environmental selection). The settings file will be ignored if no environmental selection is specified.

Example: `envifile_2`

```

2 3
0.0 0.5 1.0
0.5 0.5 0.5
1.0 0.5 0.0
0.5 0.5 0.5

```

Outputfiles

The simulation output is split into different files, most of them can be read with R. These are `file.main`, `file.haplo`, `file.stats`, `file.sel` and `file.stain`. At the end of a simulation run, an overview about the specified settings will be written to `file.log`.

`file.main` stores rows with marker genotypes (0, 1 or 2 *a* alleles) for each replicate, generation, deme and individual (every 10 generations as default, but this can be changed via the command-line). Admixture proportion (*q*) and intra-source ancestry (*het*), fitness, and the number of junctions for each individual are given at the beginning of each line.

`file.haplo` is similar to `file.main` but with two lines for each individual (indicating gametes from the maternal or paternal parent; note that all other info is simply copied, thus the number of junctions etc. is for the individual, not the gametes). `file.haplo` is useful for example for calculating LD. Instead of the counts for *a* alleles in `file.main`, allele states (i.e., *a* or *b*; or *a*, *A*, *b* or *B* with staining) are output.

`file.stats` provides a summary for several measurements within demes, and is printed for each replicate, deme and metapopulation, every generation. For each deme, the following is output: realized deme size *N*, admixture proportion *q* (mean and sd), inter-source ancestry *het* (mean and sd), fitness (mean and sd), the number of junctions (mean and sd), and deme-specific- F_{IS} . For the metapopulation, the following is output: realized metapopulation size *N*, *q* (mean), *het* (mean), genome-wide F_{ST} and metapopulation- F_{IS} . F_{IS} and F_{ST} are directly calculated from *q* and *het*. $F_{ST} = (H_T - H_S)/H_T$, where H_T is the expected heterozygosity in the metapopulation, and H_S is the expected heterozygosity within demes, averaged over all demes. The expected heterozygosity is given by $2\bar{q}(1 - \bar{q})$, where \bar{q} is the average *q* over individuals of the deme or the metapopulation. $F_{IS} = (H_E - H_O)/H_E$, where H_E is the expected heterozygosity calculated from \bar{q} as above, and H_O is the observed heterozygosity, directly taken from average *het*. The difference between deme-specific- and metapopulation- F_{IS} is that the first averages over individuals within a deme, the latter over all individuals in the metapopulation.

`file.sel` is only output if loci under selection were simulated, with one line for each replicate, generation, deme and individual (format `r*_g*_d*_i*`; every 10 generations as default, but this can be changed via the command-line). Several lines of information for each selected locus (echoed from the input files) are printed, followed by the genotypes for each locus under selection (0, 1 or 2 *a* alleles).

`file.stain` is only output if staining was specified, one line for each replicate, generation and deme (writing to this file starts not earlier than the first generation of staining; every 10 generations as default, as above). For each marker locus, the deme's allele frequency of dye *A* or *B* are given. 'Staining' corresponds to replacing all *a* or *b* alleles in the parental source populations by alleles *A* or *B* at a certain time point during the simulation. These new alleles ('dyes') will then spread through the demes by migration and gene flow. Their diffusion speed will be influenced by the specified genetic and spatial architectures, and can provide clues about barrier strength. Computation of fitness, *q*, *het*, and the number of junctions will not be impaired by the dyes. (Note that the 'dyes' *A* and *B* are disparate from the incompatible DMI alleles.)