DifferInt

Compositional differentiation among populations at three levels of genetic integration

User Manual

May 2013

Elizabeth M. Gillet

Abt. Forstgenetik und Forstpflanzenzüchtung
Georg-August-Universität Göttingen,
Büsgenweg 2, 37077 Göttingen, Germany
http://www.uni-goettingen.de/forstgenetik/differint
https://projects.gwdg.de/projects/differint
Email: differint@uni-goettingen.de

Gillet, E.M. (2013) *DifferInt:* Compositional differentiation among populations at three levels of genetic integration. *Molecular Ecology Resources* 13, 953-964. http://dx.doi.org/10.1111/1755-0998.12145

Contents

1 Introduction

The computer program *DifferInt* is described in the publication

> Gillet, E.M. (2013) *DifferInt:* Compositional differentiation among populations at three levels of genetic integration. *Molecular Ecology Resources* 13, 953-964. `http://dx.doi.org/10.1111/1755-0998.12145`

It calculates four measures of compositional differentiation among populations for their genetic types at three levels of genetic integration: the gene pool level, the mean single-locus level and the multilocus level. Genetic types refer to the single-locus or multilocus genotypes of diploid individuals in one or more populations or population samples at a given set of codominantly expressed marker loci.

1.1 Minimality principle

All four measures are based on the minimality principle, which states that the distance between two populations equals the minimum extent to which one population must be altered in order to make it match the other based on two tiers of dissimilarity:

- dissimilarity between two individuals as the proportion of genes in the genetic type of one individual that must be changed in order to make this individual match the other, termed the elementary genic difference d,
- dissimilarity between two populations as the minimum extent to which the frequencies of genetic types in one population must be altered in order to make this population match a second.

1.2 Pairwise genetic distance

Considering only the frequencies $(p_1, \ldots, p_n)$ and $(q_1, \ldots, q_n)$ of the genetic types in two populations $\mathcal{P}$ and $\mathcal{Q}$ at any given level, the absolute genetic

distance

$$d_0(\mathcal{P}, \mathcal{Q}) = \frac{1}{2} \sum_i |p_i - q_i|$$

measures the minimum proportion of genetic types in either one of the two populations that must be shifted to a different type in order to make this population match the other.

The minimality principle implies that genetic types showing a frequency excess in the one population should be shifted to frequency-deficient genetic types that differ in as few genes as possible, that is, minimize the elementary genic difference d. Simultaneous minimization of the extent of alteration with respect to both frequency differences and elementary genic differences yields the measure of pairwise population distance Δ. Its calculation requires application of a linear optimization algorithm. This approach and its implications for the measurement of gene associations is described in the open-access publication by Gillet & Gregorius (2008).

1.3 Genetic differentiation

The pairwise genetic distances Δ and d_0 are generalized to differentiation measures for higher numbers of populations in two ways:

- Complementary differentiation as the mean contribution of genetic types by each population to the genetic types represented in the entirety of all populations,
- Dispersive differentiation as the mean distance of the genetic types contributed by each population from the contributions of each of the others.

DifferInt calculates the following differentiation measures (see Tab. 1), all of which were developed within our working group:

- complementary genetic differentiation Δ_{SD} that considers elementary genic differences d, together with the contributions $\Delta_{SD}(j)$ of each population $\mathcal{P}(j)$,

	Pairwise distance	Complementary differentiation	Dispersive differentiation
Considering elementary genic differences	Δ	Δ_{SD}	$\bar{\Delta}$
Neglecting elementary genic differences	d_0	δ_{SD}	$\bar{\delta}$

Table 1: Measures of pairwise genetic distance between two populations and the two corresponding measures of genetic differentiation among populations. All measures are calculated by *DifferInt/* at three levels of genetic integration: gene-pool, single-locus genotypes, and multilocus genotypes.

- a new measure of dispersive differentiation $\bar{\Delta}$ that considers elementary genic differences d, together with the contributions $\bar{\Delta}(j)$ of each population $\mathcal{P}(j)$
- complementary differentiation δ_{SD} that neglects elementary genic differences, together with the contributions $\delta_{SD}(j)$ of the $\mathcal{P}(j)$th population
- a new measure of dispersive differentiation $\bar{\delta}$ that neglects elementary genic differences, together with the contributions $\bar{\delta}_{SD}(j)$ of the $\mathcal{P}(j)$th population.

1.4 Levels of genetic integration

In order to assess the effects of gene association, both homologous (*i.e.*, intralocus) and non-homologous (*i.e.*, interlocus), on genetic differentiation, all measures are calculated at three levels of genetic integration: the gene-pool, single-locus genotypes, and multilocus genotypes. A special property of the genetic distance Δ is that it cannot decrease from one level of integration to the next highest (see Gillet & Gregorius (2008) for mathematical proof; also see App. A). It follows that Δ_{SD} and $\bar{\Delta}$ also cannot decrease from one level to a higher level.

This property allows inference on the similarity of forms of association within pairs of populations, since it remains equal if the genotypes in the two pop-

ulations are composed by the same form of gene association. $\bar{\Delta}$ also remains equal if the genotypes in the two populations are composed by the same form of gene association. This is not true of Δ_{SD}, since the mixture of populations with the same form of association adds new association to the complement of a population, e.g. the increase in homozygosity in mixtures of populations that show Hardy-Weinberg-Proportions (Wahlund-Effect).

1.5 Snail diagrams

Subpopulation differentiation is illustrated by so-called snail diagrams. Each population is represented by a sector of a pie-like chart. The radius of the sector equals the contribution of one of the populations to differentiation (*i.e.*, $\Delta_{SD}(j)$, $\bar{\Delta}(j)$, $\delta_{SD}(j)$, or $\bar{\delta}(j)$), while the angle of the sector is proportional to the relative size of this population among all populations. The sectors are arranged by radius, with the largest contribution placed to the right of 90° and the sector with the second-largest radius to its right, etc., yielding a figure that is reminiscent of the shell of a snail. A dotted circle of radius equal to the weighted mean of the sector radii marks the population differentiation (*i.e.*, Δ_{SD}, $\bar{\Delta}$, δ_{SD}, or $\bar{\delta}$). The snails are encoded as *fig*-files for vector graphic software *WinFIG* (`http://www.winfig.com`) for Windows or *xfig* (`http://www.xfig.org`) for Linux. Optionally, snails are displayed on the screen. Examples are shown in Figs. 4-7.

1.6 Covariation of differentiation

A measure of covariation C expresses the degree of correspondence between the contributions of the single populations to differentiation over two levels of integration (see Sec. 7.4).

1.7 Statistical analysis

Statistical significance is estimated using two kinds of permutation analysis (see Sec. 7.5):

- Permutation of alleles among individuals within populations tests the homologous association of alleles into single-locus genotypes as well as

the non-homologous association of single-locus genotypes into multilocus genotypes.

- Permutation of multilocus genotypes among populations tests the association, or partitioning, of genetic types among the populations.

Results are expressed as confidence intervals of chosen percentage of permutation values, minimum and maximum permutation values, and P-values.

2 Citing *DifferInt*

Gillet, E.M. (2013) *DifferInt*: Compositional differentiation of populations at three levels of genetic integration. *Molecular Ecology Resources* 13, 953-964. `http://dx.doi.org/10.1111/1755-0998.12145`

3 Installation

3.1 Installing *DifferInt*

Download the current version of *DifferInt* from either of the following websites:
`http://www.uni-goettingen.de/forstgenetik/differint`
`https://projects.gwdg.de/projects/differint`

- Windows (32-bit)

  Download the file `DifferInt-W32.zip` to the desired directory and unzip it. A new directory named `DifferInt-W32` is created in the current directory that contains the following files

`DifferInt.exe`	Program executable
`exa1.txt`	First example input file
`exa2.txt`	Second example input file
`Gillet-Gregorius-BMCGenetics-2008.pdf`	Original publication
`metapop1.dat`	Example M^A in Gillet (2013)
`metapop2.dat`	Example M^B in Gillet (2013)
`metapop3.dat`	Example M^C in Gillet (2013)
`metapop4.dat`	Example M^D in Gillet (2013)
`GNU-General-Public-License.txt`	GNU General Public License
`README`	Instructions, notes and bug fixes

- Linux

  Download `DifferInt-linux.tar` to the desired directory and unpack it, e.g. by opening terminal window in this directory and typing

  `> tar -xvf DifferInt-linux.tar`

  A new directory named `DifferInt-linux` is created in the current directory that contains the following files

`DifferInt`	Program executable
`exa1.txt`	First example input file
`exa2.txt`	Second example input file
`Gillet-Gregorius-BMCGenetics-2008.pdf`	Original publication
`GNU-General-Public-License.txt`	GNU General Public License
`metapop1.dat`	Example M^A in Gillet (2013)
`metapop2.dat`	Example M^B in Gillet (2013)
`metapop3.dat`	Example M^C in Gillet (2013)
`metapop4.dat`	Example M^D in Gillet (2013)
`README`	Instructions, notes and bug fixes

3.2 Installing plot software

If selected in the input menu, *DifferInt* produces snail diagrams (see Figs.4-7) as files with names ending in `.fig`. To display, print, and alter these diagrams, additional software must be installed on the computer (`xxx` stands for the current version number):

- For Windows, *DifferInt* displays snails by calling the shareware *WinFIG*. Install *WinFIG* and *WinFIG_fontspack_1.0.zip* as described.

- For Linux, *DifferInt* displays snails by calling the open-access software *xfig* and *transfig*. If they are not included in your Linux distribution, install them from `http://www.xfig.org` as described. Installation in the directory `/usr/local/lib` requires root authorization.

4 Input file

4.1 Construction of input file

A file containing the genetic data is prepared in advance using a text editor or spreadsheet software. Consecutive columns of data may be separated by any number of blanks, a single comma, or a single TAB. When finished, save or export the data as a text-only(!) file (i.e., without formatting) named, for example, `input.txt`. Place it in the same directory as the executable.

Line-by-line description of input files:

* Line 1: At least one character "`#`" followed by name of data
* Line 2: Number of populations
* Line 3: Number of gene loci
* Line 4: "`yes`", if each line in the list of genetic types is to begin with the count of individuals that possess this genetic type; "`no`" otherwise
* Line 5: Empty line (optional)
* Line 6: At least one character "`#`" followed by name of population 1 (optional)
* Lines 7ff: Population 1
  - If line 4 contains "`yes`", each subsequent line contains a number of individuals (integer) of the same genetic type followed by specification of this (multilocus) genotype as a list of its two alleles at gene locus 1, then gene locus 2, etc. Alleles are designated by up to 3 numerical digits (integers). See example input file `exa1.txt` in Fig. 1.

Figure 1: Example input file `exa1.txt`

```
#Example1
3    #No. populations
2    #No. loci
yes  #Each genotype is preceded by number of individuals (yes|no)

#Population 1
2    1 1   1 1
2    1 1   1 2
3    1 1   2 2
3    1 2   1 1
2    1 2   1 2
3    1 2   2 2
3    2 2   1 1
2    2 2   1 2
2    3 3   2 3

#Population 2
2    1 1   1 1
1    1 1   1 2
1    1 1   2 2
1    1 2   1 1
1    1 2   1 2
1    1 2   2 2
1    2 2   1 1
1    2 2   1 2
1    2 2   2 2

#Population 3
2    1 1   1 1
2    1 1   3 3
2    1 1   4 4
2    1 3   1 1
2    1 3   1 2
2    1 4   4 4
2    2 2   1 2
2    2 3   2 3
2    4 4   4 4
```

Figure 2: Example input file `exa2.txt` – same data as in Fig. 1

```
#Example1
3    #No. populations
2    #No. loci
no  #Each genotype is preceded by number of individuals (yes|no)

#Population 1
1 1   1 1
1 1   1 1
1 1   1 2
1 1   1 2
1 1   2 2
1 1   2 2
1 1   2 2
1 2   1 1
1 2   1 1
1 2   1 1
1 2   1 2
1 2   1 2
1 2   2 2
1 2   2 2
1 2   2 2
2 2   1 1
2 2   1 1
2 2   1 1
2 2   1 2
2 2   1 2
3 3   2 3
3 3   2 3

#Population 2
1 1   1 1
1 1   1 1
1 1   1 2
1 1   2 2
1 2   1 1
1 2   1 2
1 2   2 2
2 2   1 1
2 2   1 2
2 2   2 2
```

```
#Population 3
1 1   1 1
1 1   1 1
1 1   3 3
1 1   3 3
1 1   4 4
1 1   4 4
1 3   1 1
1 3   1 1
1 3   1 2
1 3   1 2
1 4   4 4
1 4   4 4
2 2   1 2
2 2   1 2
2 3   2 3
2 3   2 3
4 4   4 4
4 4   4 4
```

[input file continued to the right] ↗

  - If line 4 contains "`no`", each subsequent line contains only the (multilocus) genotype of a single individual, as specified in the previous paragraph. Thus two individuals possessing the same genetic type occupy two lines of input. See example input file `exa2.txt` in Fig. 2.

* Empty line

* For each subsequent population, repeat previous lines beginning with line 6

* End file with one empty line (or one line filled with any number of blanks)

4.2 Example of input files

The two example input files `exa1.txt` in Fig. 1 and `exa2.txt` in Fig. 2 are equivalent in that they show the same data. In `exa1.txt`, individuals possessing the same genotype are collected into one line headed by their number. In `exa2.txt`, each individual occupies a line of its own with its genotype.

5 Program execution

Execution of *DifferInt* is initiated by one of three methods:

5.1 Menu-driven execution

5.1.1 Start program

In a file manager, start execution by double-clicking on

`DifferInt.exe` (for Windows)

`DifferInt` (for Linux)

or in a terminal window, start execution by typing

> `DifferInt` (for Windows)

> `./DifferInt` (for Linux)

!!!!!!!! IMPORTANT NOTE FOR LINUX USERS !!!!!!!!!!

On some computers, warnings of the following type appear

```
Warning: translation table syntax error: Unknown keysym name:
osfActivate
Warning: ... found while parsing ':<Key>osfActivate:
ManagerParentActivate()'
```

As explained by Helmut Michels (2011) in the DISLIN Discussion Group, when a program that uses DISLIN widgets is executed on a recent or older Linux system than the one on which it was created, warnings of this type may appear. The reason is that the location of the file `XKeysymDB` used by the widgets can differ between Linux systems. A workaround is to define the environment variable `XKEYSYMDB` with the location of the `XKeysymDB` file. Thus if

```
> find /usr/ -name XKeysymDB
```
or
```
> locate XKeysymDB
```

returns `location`, then enter

```
> export XKEYSYMDB=location
```

For example, if `location` is `/usr/share/X11/XKeysymDB`, then type

```
> export XKEYSYMDB=/usr/share/X11/XKeysymDB
```

and restart the program by typing

```
> ./DifferInt
```

5.1.2 File select window

A window opens for selection of the input file, e.g. `input.txt`. Close window by clicking `OK`.

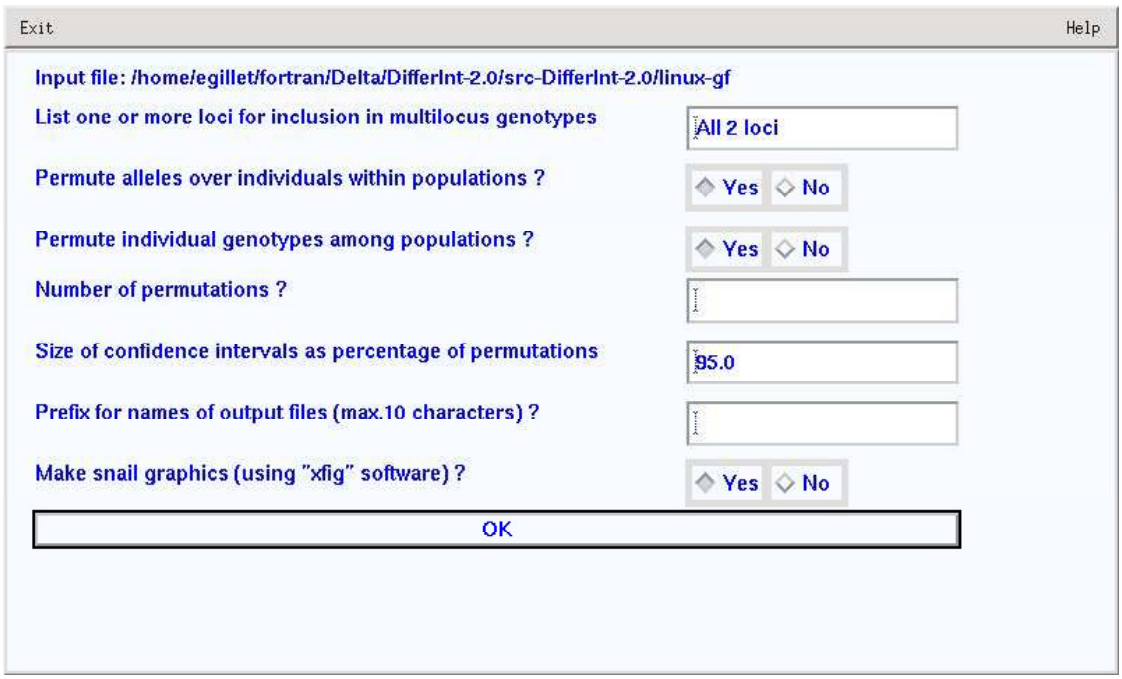


Figure 3: Interactive input window for menu-driven execution

5.1.3 Interactive input window

An interactive input window opens (see Fig. 3) requesting the following specifications:

* `List one or more loci for inclusion in multilocus genotypes`:

  To select all loci, retain preselected text "`All X loci`". Otherwise, replace text by a list of loci by number, separated by blanks or commas. The number of loci in the data is determined by a preliminary reading of the input file.

* `Permute alleles over individuals within populations ?`:

  Select "`Yes`" or "`No`" (see Sec.7.5.1).

* `Permute individual genotypes among populations ?`:

  Select "`Yes`" or "`No`" (see Sec.7.5.2).

* `Number of permutations?`:

  Insert a large integer. There is no default value.

* `Size of confidence intervals as percentage of permutations`:

  Either retain preselected value of "`95.0`" or replace by desired percentage "`X`" (with or without decimal point).
  The confidence interval is a result of permutation analysis. It is taken to span the central $X\%$ of the parameter values yielded by all permutations. Thus the $\frac{1}{2} \cdot X\%$ smallest and the $\frac{1}{2} \cdot X\%$ largest values are excluded from the confidence interval.

* `Prefix for names of output files (max.10 characters)?`:

  Insert up to 10 characters to designate output files. If, for example, "`aaa`", is entered, the names of all output files will begin with "`aaa`". There is no default prefix.

* `Make snail graphics (using WinFig software?)` (for Windows)
  or
  `Make snail graphics (using xfig software)?` (for Linux):

For each level of genetic integration, the snails rank the populations as pie pieces of decreasing radius around a central point, where the radius assigned to each population is proportional to the genetic distance of this population to the complement population formed by pooling all other populations (Gillet & Gregorius, 2008; Gregorius & Roberds, 1986).

If the answer is "`yes`", two files are prepared for the creation of snail graphics (for prefix "`aaa`") that serve as input to the vector graphics program *xfig* (for Linux) or *WinFig* (for Windows) (see Sec. 3.2).

`aaa.Delta.schnecken.fig`: These snails represent the genetic distances $\Delta_{SD}(j)$, which consider the elementary genic differences between genetic types

`aaa.d0.schnecken.fig`: These snails represent the genetic distances D_j, which neglect genic differences

If the answer is "`no`", the two `.fig`-files are produced but not displayed. They can be called up later and modified by one of the vector graphics programs *WinFIG* for Windows or *xfig* for Linux (see Sec. 3.2).

5.2 Keyboard-driven execution

This form of execution is initiated by starting *DifferInt* with the option `-nomenu`, *i.e.*,

> `DifferInt -nomenu` (for Windows)

> `./DifferInt -nomenu` (for Linux)

A line then appears asking for the name of the input file:

> `Input name of input file :`

Enter name of file and press <Return>.

A listing of the program's heading appears, followed by a series of questions to be answered over the keyboard. The questions are the same as for menu-driven execution (see Sec. 5.1.3).

5.3 Batch-type execution

A file (e.g. `batch.txt`) can be prepared in advance with a text editor that contains the answers to all of the questions posed under keyboard-driven execution, just as if they were being typed directly into the terminal window. *DifferInt* is started by typing into a terminal window (without a minus-sign)

> `DifferInt batch.txt` (for Windows)

> `./DifferInt batch.txt` (for Linux)

6 Command-line arguments

When execution is initiated from a terminal window, up to three arguments can be added on the command line (brackets `[]` signify that inclusion is optional)

> `DifferInt [arg1 [arg2 [arg3] ] ]` (for Windows)

> `./DifferInt [arg1 [arg2 [arg3] ] ]` (for Linux)

The arguments can be of the following forms:

- If `arg1` does not begin with a minus sign, it is taken to be the name of a batch file containing all specifications (e.g. the file `batch.txt` discussed in Sec. 5.3). Additional command-line arguments may follow, but this must be the first.
- `-nomenu`

  Name of input file and program specifications are to be requested through the keyboard (see Sec. 5.2).
- `-ncols=X`

  The positive integer `X` specifies the maximum number of columns that are printed on a line in the output file. If the results contain more than `X` columns, the line is wrapped. By default, `-ncols=10`.

7 Calculations

DifferInt calculates the following measures at all three levels of genetic integration, then determines confidence intervals and P-values by two forms of permutation analysis.

7.1 Levels of genetic integration

Genetic integration signifies the form of association of genes to single-locus genotypes to multilocus genotypes. *DifferInt* considers three levels of genetic integration:

7.1.1 Gene-pool

The lowest level of genetic integration is the gene-pool. The gene-type of each individual gene is characterized by the gene locus at which it is located and by its allelic state. Assuming that the degree of ploidy is the same at all loci, the relative frequencies of the gene-types in the gene-pool of a population equal $(1/L) \cdot p_{i;l}$, where L is the number of loci L and $p_{i;l}$ is the relative frequency of the i-th allele at the l-th gene locus in the population ($\sum_i p_{i;l} = 1$ and $\sum_{i,l}(1/L) \cdot p_{i;l} = 1$).

7.1.2 Single-locus genotypes

In polyploid species, more than one gene is present at the same gene locus. In diploid species, two genes are present at all (autosomal) loci. Thus the genetic type of each individual at a given locus is characterized by the allelic types of its genes at this locus, *i.e.*, its single-locus genotype. Within the population, the form of association of the alleles in the gene-pool into single-locus genotypes determines the distribution of single-locus genotypes. Random association yields Hardy-Weinberg proportions.

7.1.3 Multilocus genotypes

When considering more than one gene locus simultaneously, the genetic type of an individual is characterized by the list of its single-locus genotypes at each of these loci, *i.e.*, its multilocus genotype. Within the population, the form of association of the single-locus genotypes of the individuals into multilocus genotypes determines the distribution of multilocus genotypes. Random association appears as gametic equilibrium.

7.2 Measures considering elementary genic differences

At each level of the three levels of genetic integration, *DifferInt* calculates the following measures. These measures are based not only on relative frequencies of the genetic types of the individuals in the populations but also on the degree to which the genetic types themselves differ from each other.

7.2.1 Elementary genic difference d between genetic types

The genic difference between two genetic types at the same level of integration is basically determined by the number of their individual genes that differ in allelic type. If the numbers of copies of the i-th allele at the l-th gene locus are denoted by $n_{i;l}$ and $m_{i;l}$, respectively, then the two genetic types differ by $\sum_{i,l} |n_{i;l} - m_{i;l}|$ gene-type copies. This sum is maximal, equaling two times the total number K of individual genes represented in each object, if the objects share no gene-types (and thus differ completely). Since $\sum_{i,l} n_{i;l} = \sum_{i,l} m_{i;l} = K$ holds, division of $\sum_{i,l} |n_{i;l} - m_{i;l}|$ by $2K$ yields a measure of genic difference that is bounded between zero and one. This measure of elementary genic difference is applicable to all levels of integration.

7.2.2 Pairwise genetic distance Δ between populations

The pairwise genetic distance

$$\Delta(\mathcal{P}, \mathcal{Q})$$

equals the minimum degree to which the frequency distribution of the genetic types within one of the populations, say $\mathcal{P}$, must be transformed in order to make it match the distribution in the other, say $\mathcal{Q}$, where the degree of transformation is based on the elementary genic differences between genetic types (Gregorius *et al.*, 2003; Gillet *et al.*, 2004).

A transformation s consists of a set of frequency shifts within population $\mathcal{P}$ from types that are more frequent in population $\mathcal{P}$ than in $\mathcal{Q}$ to types that are less frequent in $\mathcal{P}$ than in $\mathcal{Q}$ for the purpose of making population $\mathcal{P}$ match population $\mathcal{Q}$. If the frequency p_a of type a in $\mathcal{P}$ exceeds the frequency q_a of this type in $\mathcal{Q}$, then the excess $p_a - q_a$ must be shifted to types deficient in $\mathcal{P}$, such that $\sum_b s(a,b) = p_a - q_a = p_a - \min\{p_a, q_a\}$. The shift process is continued for all types with a frequency excess in $\mathcal{P}$ until the frequencies of all types in $\mathcal{P}$ match those in $\mathcal{Q}$. Denoting s as the transformation, *i.e.*, the totality of frequency shifts, and weighting each shift $s(a,b)$ by the elemtary genic difference $d(a,b)$ between the respective genetic types yields

$$\Delta(s) = \sum_{a,b} s(a,b) \cdot d(a,b)$$

$\Delta(s)$ can be considered as the "cost" of transformation s.

It is easy to see that at least one such transformation exists. Since there may be more than one, Δ is taken to be the minimum value of $\Delta(S)$ over all permissible transformations S, *i.e.*,

$$\Delta(\mathcal{P}, \mathcal{Q}) = \min_S \Delta(S)$$

Gregorius *et al.* (2003) and Gillet *et al.* (2004) show that finding a shift transformation s that minimizes Δ(s) is equivalent to solving the cost-minimizing *Transportation Problem* (Hitchcock, 1941) of operations research. Linear programming methods for its solution are implemented in the computer program *DeltaS* (`http://www.uni-goettingen.de/de/95605.html`).

See App. A for an explanation of special relationships between values of the genetic difference Δ when calculated at different levels of genetic integration (gene-pool, mean single-locus genotypes, and multilocus genotypes).

DifferInt outputs lower triangular matrices containing the pairwise distances Δ for all pairs of populations. These distance matrices can be input into

other programs that make dendograms or choropleths, for example.

7.2.3 Complementary subpopulation differentiation Δ_{SD}

The complementary subpopulation differentiation

$$\Delta_{SD} = \sum_{j=1}^{n} c_j \cdot \Delta_{SD}(j)$$

(Gillet & Gregorius, 2008) equals the weighted arithmetic mean of the genetic distances

$$\Delta_{SD}(j) = \Delta(\mathcal{P}(j), P^c(j))$$

of each subpopulation $\mathcal{P}(j) = (p_1(j), \ldots, p_n(j))$ to its complement population $\mathcal{P}^c(j) = (p_1^c(j), \ldots, p_n^c(j))$ that is formed by pooling all other populations, where $p_i^c(j) = \sum_{k \neq j} p_i(k)/(1 - c_j)$ and the weight c_j is the relative size of subpopulation $\mathcal{P}(j)$. In *DifferInt*, subpopulation sizes are disregarded, yielding

$$c_j \equiv 1/n \quad \text{and} \quad p_i^c(j) = \sum_{k \neq j} p_i(k) \cdot n/(n-1)$$

The name of this measure indicates that it was conceived for the measurement of differentiation among subpopulations belonging to a larger population, but it is applicable to any set of populations. Δ_{SD} ranges between 0 and 1. Δ_{SD} equals 1 if all populations are genetically disjoint, *i.e.*, no two populations share the same genetic types. Δ_{SD} equals 0 if all populations show identical frequency distributions for the genetic types.

7.2.4 Dispersive subpopulation differentiation $\bar{\Delta}$

The dispersive subpopulation differentiation

$$\bar{\Delta} = \sum_{j=1}^{n} c_j \cdot \bar{\Delta}(j)$$

equals the weighted arithmetic mean of the mean genetic distance Δ of each population $\mathcal{P}(j)$ to each of the others, *i.e.*,

$$\bar{\Delta}(j) = \frac{1}{1 - c_j} \cdot \sum_{k \neq j} c_k \cdot \Delta(\mathcal{P}(j), \mathcal{P}(k))$$

of each subpopulation $\mathcal{P}(j) = (p_1(j), \ldots, p_n(j))$ to each of the others and the weight c_j is the relative size of subpopulation $\mathcal{P}(j)$. In *DifferInt*, subpopulation sizes are disregarded, yielding

$$c_j \equiv 1/n \quad \text{and} \quad p_i^c(j) = \sum_{k \neq j} p_i(k) \cdot n/(n-1)$$

The name of this measure indicates that it is an average. $\bar{\Delta}$ ranges between 0 and 1. $\bar{\Delta}$ equals 1 if all populations are genetically disjoint, *i.e.*, no two populations share the same genetic types. $\bar{\Delta}$ equals 0 if all populations show identical frequency distributions for the genetic types.

7.3 Measures neglecting elementary genic differences

At each level of the three levels of genetic integration, *DifferInt* also calculates the following measures. These are equivalent to the measures in the previous sections with the exception that the differences between genetic types are neglected. By replacing the elementary genic difference between genetic types by the discrete difference, the measures given below are based only on relative frequencies of the genetic types of the individuals in the populations.

7.3.1 Discrete difference between genetic types

The discrete difference between two genetic types at the same level of genetic integration equals 0 if the types are identical over all alleles at all of the loci under consideration and 1 if the types differ by at least one allele at one locus.

7.3.2 Pairwise genetic distance d_0 between two populations

The pairwise genetic distance

$$d_0(\mathcal{P}, \mathcal{Q}) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i|$$

(Gregorius, 1974) between populations $\mathcal{P}$ of frequency distribution $\mathcal{P} = (p_1, \ldots, p_n)$ and $\mathcal{Q}$ of frequency distribution $\mathcal{Q} = (q_1, \ldots, q_n)$ equals the minimum degree to which the frequency distribution of the genetic types within one of the populations must be transformed in order to make it match the distribution of types in the other population (Gregorius *et al.*, 2003; Gillet *et al.*, 2004; Gregorius, 1987, 1984).

Δ equals d_0 if the "cost" of transformation for Δ is taken to be the discrete difference. Values of d_0 range between 0 and 1. d_0 equals 1 if the two populations are genetically disjoint, *i.e.*, do not share any genetic type. d_0 equals 0 if $\mathcal{P} = Q$.

DifferInt outputs lower triangular matrices containing the pairwise distances d_0 for all pairs of populations. These distance matrices can be input into other programs that make dendograms or choropleths, for example.

7.3.3 Complementary subpopulation differentiation δ_{SD}

The complementary subpopulation differentation

$$\delta_{SD} = \sum_{j=1}^{n} c_j \cdot D(j)$$

(Gregorius, 1996; Gregorius & Roberds, 1986) equals the weighted arithmetic mean of the genetic distances

$$\delta_{SD}(j) = D_j = d_0(\mathcal{P}(j), \bar{P}(j)) = \frac{1}{2} \sum_{i=1}^{n} |p_i(j) - \bar{p}_i(j)|$$

of each subpopulation $\mathcal{P}(j) = (p_1(j), \ldots, p_n(j))$ to its complement population $\bar{P}(j) = (\bar{p}_1(j), \ldots, \bar{p}_n(j))$ that is formed by pooling all other populations,

where $\bar{p}_i(j) = \sum_{k \neq j} p_i(k)/(1 - c_j)$ and the weight c_j is the relative size of subpopulation $\mathcal{P}(j)$. In *DifferInt*, subpopulation sizes are disregarded, yielding

$$c_j \equiv 1/n \quad \text{and} \quad \bar{p}_i(j) = \sum_{k \neq j} p_i(k) \cdot n/(n-1)$$

Δ_{SD} equals δ_{SD} if the "cost" of transformation for the pairwise difference Δ is taken to be the discrete difference. Values of δ_{SD} range between 0 and 1. δ_{SD} equals 1 if all populations are genetically disjoint, *i.e.*, no two populations share the same genetic types. δ_{SD} equals 0 if all populations show identical frequency distributions for the genetic types.

7.3.4 Dispersive subpopulation differentiation $\bar{\delta}$

The dispersive subpopulation differentiation

$$\bar{\delta} = \sum_{j=1}^{n} c_j \cdot \bar{\delta}(j)$$

equals the weighted arithmetic mean of the mean genetic distance d_0 of each population $\mathcal{P}(j)$ to each of the others, *i.e.*,

$$\bar{\delta}(j) = \frac{1}{1 - c_j} \cdot \sum_{k \neq j} c_k \cdot d_0(\mathcal{P}(j), \mathcal{P}(k))$$

of each subpopulation $\mathcal{P}(j) = (p_1(j), \ldots, p_n(j))$ to each of the others and the weight c_j is the relative size of subpopulation $\mathcal{P}(j)$. In this version of *DifferInt*, subpopulation sizes are considered equal. $\bar{\delta}$ ranges between 0 and 1. $\bar{\delta}$ equals 1 if all populations are genetically disjoint, *i.e.*, no two populations share the same genetic types. $\bar{\delta}$ equals 0 if all populations show identical frequency distributions for the genetic types.

7.4 Covariation of differentiation between integration levels

The degree of correspondence between differentiation indices from two levels of integration can be determined by a measure of covariation. As was pointed

out in Gregorius *et al.* (2007), a suitable measure of covariation is

$$C = \frac{\sum_{i<k}(X_i - X_k)(Y_i - Y_k)}{\sum_{i<k}|(X_i - X_k)(Y_i - Y_k)|}$$

where the variables X_j and Y_j refer to the contribution of population $\mathcal{P}(j)$ at two different levels of integration and X_k and Y_k to the respective contributions of a different population k. In the case of complementary differentiation Δ_{SD} and δ_{SD}, X_j and Y_j refer to $\Delta_{SD}(j)$ and $\delta_{SD}(j)$, respectively, at two levels of integration. In the case of dispersive differentiation $\bar{\Delta}$ and $\bar{\delta}$, X_j and Y_j refer to $\bar{\Delta}(j)$ and $\bar{\delta}(j)$, respectively, at two levels. C varies between -1 and $+1$ such that $C = 1$ for strictly positive and $C = -1$ for strictly negative covariation. It is undefined in the practically irrelevant case where a non-zero difference for one variable implies equality for the other.

7.5 Statistical analysis

The significance of the observed measures of distance and differentiation at each level of integration is estimated by two kinds of permutation analysis.

7.5.1 Testing random association of genes to genotypes within populations

As stated by Gillet & Gregorius (2008), "Any increase of genetic differentiation among populations at higher levels of genetic integration is due to forces of association of genes that differ among populations. It is thus of basic interest to know whether the differentiation observed at a level of integration can be explained by random combination of genes (e.g. into diploid genotypes or haplotypes) or whether directed forces of combination must be assumed. This requires an analysis that is conditional on the gene-pool of each population, the number of populations, and the population sizes. The effects of chance can be assessed by permuting the genes within each population, such that all homologous and non-homologous combinations of genes (alleles) into (haploid, diploid or polyploid) genotypes have equal probability."

For each permutation, at each gene locus the genes present within each population are randomly reassigned to the individuals, creating new distributions

of single- and multilocus genotypes within populations and, accordingly, new realizations of the associations of the genetic types at one level of genetic integration into the genetic types at the next higher level (*i.e.*, gene-pool to single-locus genotypes, single-locus genotypes to multilocus genotypes).

The values of all measures at each integration level are calculated for each permutation. After all permutations have been completed, the P-value for each measure is determined as the proportion of all permutations that yielded values greater than or equal to the observed value. For interpretation of the results, both very small P-values (0.05) and very large P-values (0.95) are of interest. *DifferInt* prints the following results from the permutations for each measure (see Tab.2):

- observed differentiation
- minimum value of differentiation over all permutations
- lower and upper bounds of the confidence interval as the central $X\%$ of the values from all permutations (where "`X`" is the chosen "`Size of confidence intervals as percentage of permutation`")
- maximum value of differentiation over all permutations
- P-value for the observation

It may occur that the hypothesis is rejected at one level of integration but not another. This can provide clues to the cause of the non-random association.

7.5.2 Testing random assignment of genetic types to populations

Populations are differentiated when genetic types are unevenly distributed over populations. To test whether genetic types to populations could have been randomly assigned to the populations, the (multilocus) genotypes of all individuals in the metapopulation are randomly permuted among the populations, resulting in new assignments of genetic types to populations. The total number of individuals with each genetic type at each level of integration remains the same.

For each permutation, *DifferInt* calculates the four differentiation measures at each integration level. After completion of all permutations, the analogous

results as in Sec. 7.5.1 are listed in the output (see Tab.2). Rejection of the hypothesis at one level of integration but not another can provide clues to the cause of the non-random assignment.

8 Output files

Besides appearing in a window, results are printed to the following files, where the prefix "`aaa`" is that specified during program initiation:

* `aaa.summary.txt`
  - SUMMARY OF OBSERVATIONS AND ALL PERMUTATIONS
  - DIFFERENTIATION at levels of genetic integration GenePool, MnSLGeno, and MultGeno
  - COVARIATION OF DIFFERENTIATION between levels of integration

  > Observed values

  > Permutation analysis: Minimum, Confidence Interval, Maximum, P-Value using differentiation measures that consider elementary genic differences

  > Complementary population differentiation Δ_{SD}

  > Dispersive differentiation $\bar{\Delta}$

  > Complementary population differentiation δ_{SD}

  > Dispersive differentiation $\bar{\delta}$

* `aaa.alleles.txt`, `aaa.singlgeno.txt`, and `aaa.multigeno.txt`
  More detailed output for the each of the three levels of integration

  > Frequency distributions of the genetic types in the populations

  > Matrices of elementary genic differences between genetic types

  > Optimal shift matrices of the distance measure Δ for all comparisons between two populations and between each population and its complement population at all levels of genetic integration

* `aaa.excluded.txt`
  List of (multilocus) genotypes that are excluded from the analysis due to missing data

* `aaa.d0matrix.txt` and `aaa.Deltamatrix.txt`
  List of distance matrices for the different levels of genetic integration

* `aaa.bigdeltasd.schnecken.fig`,
  `aaa.bigdeltabar.schnecken.fig`,
  `aaa.deltasd.schnecken.fig`,
  `aaa.deltabar.schnecken.fig`
  Input files for plot software `xfig` (Linux) or `winfig` (Windows). The `fig`-files are constructed from the files of the same names ending in `.in`. If the answer to `Make snail graphics` is "`yes`", the `.fig`-files are plotted on the screen and can be exported in various formats. See Figs. 4-7.

Figure 4: Snails for the complementary differentiation measure Δ_{SD} that considers elementary genic differences. Input file `exa1.txt`.

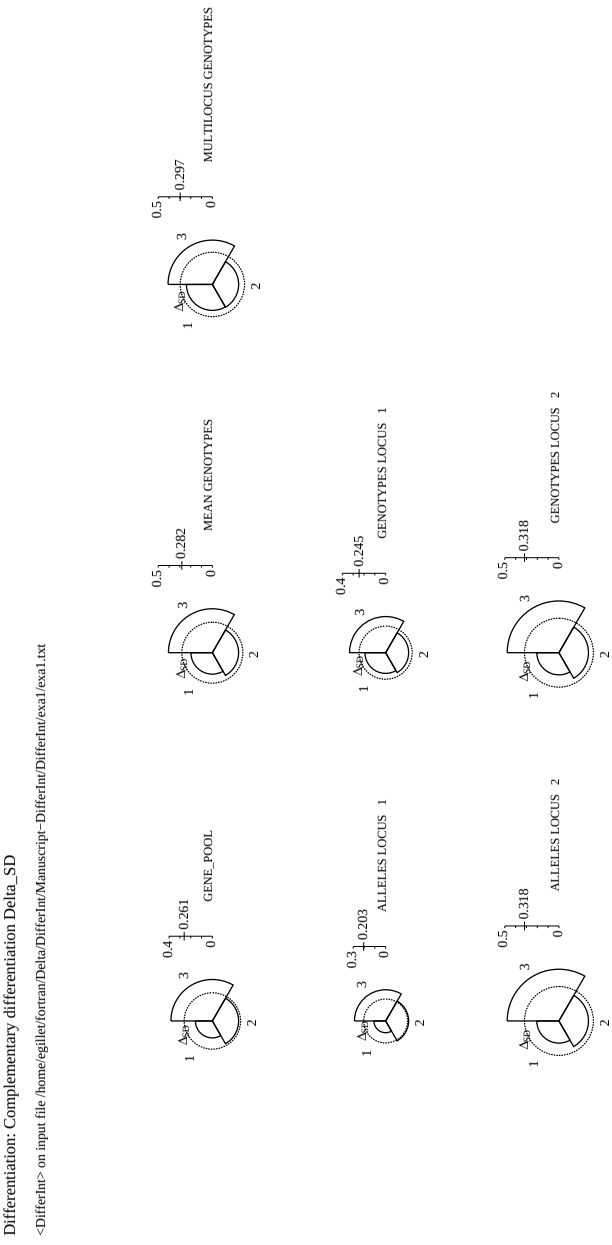


Figure 5: Snails for the dispersive differentiation measure $\bar{\Delta}$ that considers elementary genic differences. Input file `exa1.txt`.

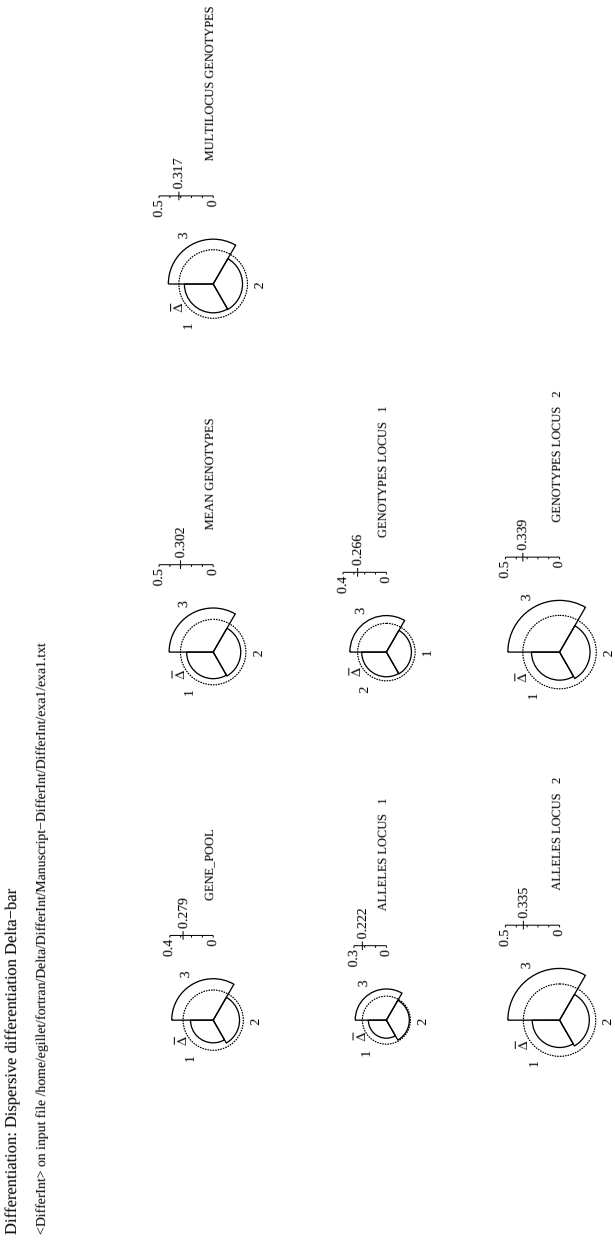


Figure 6: Snails for the complementary differentiation measure δ_{SD} that neglects elementary genic differences. Input file `exa1.txt`.

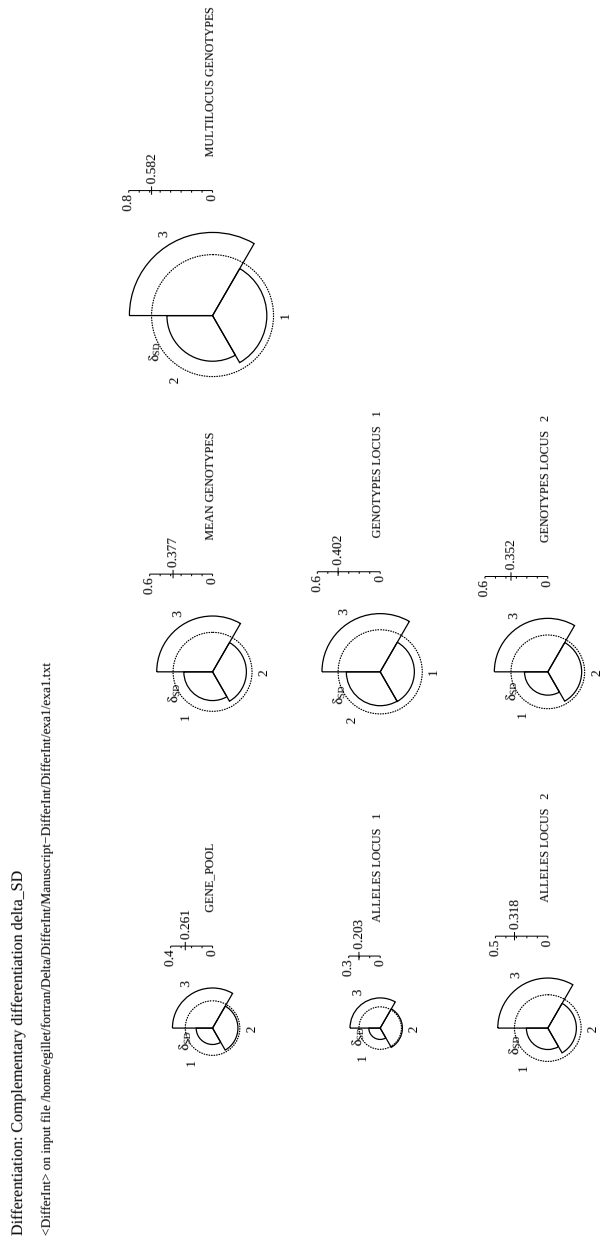
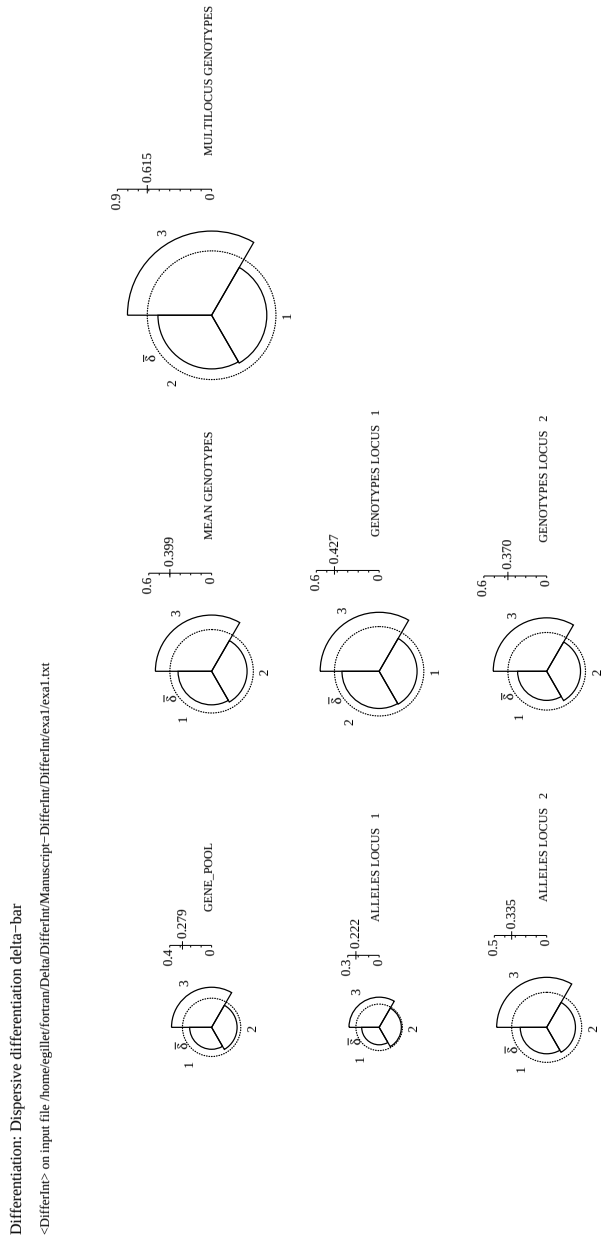


Figure 7: Snails for the dispersive differentiation measure $\overline{\delta}$ that neglects elementary genic differences. Input file `exa1.txt`.



9 Compilation

9.1 `gfortran` compiler

DifferInt is available as executables for Windows and Linux.

9.2 Windows

Compiled for Windows-XP (32-bit) using the GNU Fortran 95 compiler `gfortran` (`http://gcc.gnu.org/fortran`), version 4.8.0 20120518 (experimental) [trunk revision 187663], built for MinGW.

9.3 Linux

Compiled for Linux (32-bit) using the Fortran 95 compiler `gfortran` (`http://gcc.gnu.org/fortran`), version 4.7.1 20120723 [gcc-4_7-branch revision 189773], built on opensuse 12.2.

9.4 Implemented software

DifferInt implements subroutines from the following software:

9.4.1 *RELAX-IV*

DifferInt incorporates open-source code from the program *RELAX-IV Code for Solving the Minimum Cost Network Flow Problem* of Bertsekas & Tseng (1994) that solves the Hitchcock-Koopmans Transportation Problem known from the field of logistics by linear programming methods. This method is used to calculate the pairwise distance Δ between two populations by minimizing the extent to which the genetic types of individuals in one of the populations must be altered in order to make this population match the composition of genetic types in the other.

9.4.2 `init_random_seed()`

DifferInt applies the pseudorandom number generator `RANDOM_NUMBER` of open-source GNU Fortran (`http://gcc.gnu.org/fortran`). The random seed is initialized based on the computer system's time and thus differs for each run. The applied subroutine `init_random_seed()` is found on pp. 185ff of the manual by The `gfortran` team (1999-2013).

9.4.3 *DISLIN*

Widgets for the input and output windows are generated by routines of the *DISLIN* Scientific Plotting Software by H. Michels (`http://www.dislin.de`).

9.4.4 *WinFIG* and *xfig*

Snail diagrams are plotted using vector graphics software. For Windows, the software package *WinFIG* (`http://www.winfig.com`) must be installed on the computer. For Linux, the packages *xfig* and *transfig* (`http://www.xfig.org`) must be installed.

10 Copyright and terms of use

DifferInt - Compositional differentiation among populations at three levels of genetic integration.
©(C) 2008-2013 Elizabeth M. Gillet (Email: `egillet@gwdg.de`)
excepting code from RELAX-IV of Bertsekas & Tseng (1994)
and `init_random_seed()` of The `gfortran` team (1999-2013).

Distributed under the GNU General Public License: *DifferInt* is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, version 3. – This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. – You should have received

a copy of the GNU General Public License along with this program. If not, see http://www.gnu.org/licenses.

Send comments, suggestions, and bug reports to differint@uni-goettingen.de.

Acknowledgements: This work was funded by grant Zi 662/5-2 of the Deutsche Forschungsgemeinschaft.

References

Gillet, E.M. (2013) *DifferInt:* Compositional differentiation among populations at three levels of genetic integration. *Molecular Ecology Resources* 13, 953-964. http://dx.doi.org/10.1111/1755-0998.12145

Gillet, E.M., Gregorius, H.-R. (2008) Measuring differentiation among populations at different levels of genetic integration. BMC Genetics 9, 60. http://dx.doi.org/10.1186/1471-2156-9-60

Gillet, E.M., Gregorius, H.-R., Ziehe, M. (2004) May inclusion of trait differences in genetic cluster analysis alter our views? Forest Ecology and Management 197, 149-158. http://dx.doi.org/10.1016/j.foreco.2004.05.010

Gregorius, H-R. (1974) On the concept of genetic distance between populations based on gene frequencies. Proceedings, Joint IUFRO Meeting S.02-04, 17-22.

Gregorius, H.-R. (1984) A unique genetic distance. Biometrical Journal 26, 13-18. http://dx.doi.org/10.1002/bimj.4710260103

Gregorius, H.-R. (1987) The relationship between the concepts of genetic diversity and differentiation. Theoretical and Applied Genetics 74, 397-401. http://dx.doi.org/10.1007/BF00274724

Gregorius, H-R. (1996) Differentiation between populations and its measurement. Acta Biotheoretica 44, 23-36. http://dx.doi.org/10.1007/BF00046433

Gregorius, H.-R., Degen, B., König, A. (2007) Problems in the analysis of genetic differentiation among populations - a case study in *Quercus robur*. Silvae Genetica 56, 190-199. http://www.bfafh.de/inst2/sg-pdf/56_3-4_190.pdf

Gregorius, H.-R., Gillet, E.M., Ziehe, M. (2003) Measuring differences of trait distributions between populations. Biometrical Journal 45, 959-973. http://dx.doi.org/10.1002/bimj.200390063

Gregorius, H.-R., Roberds, J.H. (1986) Measurement of genetical differentiation among subpopulations. Theoretical and Applied Genetics 71, 826-834. http://dx.doi.org/10.1007/BF00276425

Hitchcock, F.L. (1941) Distribution of a product from several sources to numerous localities. Journal of Mathematical Physics 20, 224-230.

Software:

Bertsekas, D.P., Tseng, P. (1994) RELAX-IV: A faster version of the RELAX code for solving minimum cost flow problems. Report and FORTRAN code http://elib.zib.de/pub/Packages/mathprog/mincost/relax-4

Gillet, E.M (2003) *DeltaS*, a program to calculate the measure of pairwise distance Δ between populations. Version 2.1 (2008). http://www.uni-goettingen.de/de/95605.html

Michels, H. *DISLIN Scientific Plotting Software*. http://www.dislin.de

The gfortran team. (1999-2013) Using GNU Fortran: For GCC version 4.9.0. Free Software Foundation, Boston, MA. http://gcc.gnu.org/onlinedocs/gfortran.pdf

Schmidt, A. *WinFIG*. http://www.winfig.com

Sutanthavibul S., Smith, B.V., King, P. *et al. Xfig*. http://www.xfig.org

Beck, M., Smith, B. *Transfig*. http://www.xfig.org

A Comparing genetic distance Δ at different levels of genetic integration

For a given set of gene loci, the following always holds for the genetic distances $\Delta(\mathcal{P}, \mathcal{Q})$ between any two populations $\mathcal{P}$ and $\mathcal{Q}$ at the different levels of genetic integration:

* The mean over loci of the distances $\Delta(\mathcal{P}, \mathcal{Q})$ between the allelic frequency distributions (gene-pool) cannot be greater than the mean over loci of the distances $\Delta(\mathcal{P}, \mathcal{Q})$ between the frequency distributions of the single-locus genotypes.
* The mean over loci of the distances $\Delta(\mathcal{P}, \mathcal{Q})$ between the frequency distributions of the single-locus genotypes cannot be greater than the distance $\Delta(\mathcal{P}, \mathcal{Q})$ between the frequency distributions of the multilocus genotypes.

In some special cases, $\Delta(\mathcal{P}, \mathcal{Q})$ does not increase from a lower to a higher level of genetic integration:

Case 1: For a given gene locus, let the genotypes of the individuals in both population $\mathcal{P}$ and population $\mathcal{Q}$ show Hardy-Weinberg proportions. The distributions of the alleles may differ between the populations. Then the following measures are equal:

$\Delta(\mathcal{P}, \mathcal{Q})$ at the integration level of the single-locus genotypes

$= \Delta(\mathcal{P}, \mathcal{Q})$ at the integration level of the gene-pool (alleles)

$= d_0(\mathcal{P}, \mathcal{Q})$ at the level of the gene-pool.

The same equality holds if $\mathcal{P}$ and $\mathcal{Q}$ both show inbreeding structures with the same value of the inbreeding coefficient F (Hardy-Weinberg proportions correspond to inbreeding structures with $F = 0$.)

Case 2: For a given set of gene loci, let the multilocus genotypes of the individuals in population $\mathcal{P}$ and in population $\mathcal{Q}$ show random association of the single-locus genotypes among gene loci. The distributions of the multilocus genotypes may differ between the populations. Then the following measures are equal:

$\Delta(\mathcal{P}, \mathcal{Q})$ at the integration level of the multilocus genotypes

$= \Delta(\mathcal{P}, \mathcal{Q})$ at the integration level of the mean single-locus genotypes

Combination of Cases 1 and 2: For a given set of gene loci, let both population $\mathcal{P}$ and population $\mathcal{Q}$ show Hardy-Weinberg proportions (or equal inbreeding coefficients F) at all loci and let the single-locus genotypes be randomly associated in the multilocus genotypes of the individuals within each population. Then the following measures are equal:

$\Delta(\mathcal{P}, \mathcal{Q})$ at the integration level of the multilocus genotypes

$= \Delta(\mathcal{P}, \mathcal{Q})$ at the integration level of the mean single locus genotypes

$= d_0(\mathcal{P}, \mathcal{Q})$ at the integration level of the gene-pool.

B Example of an output file

The following pages (Tab. 2) show the output file aaa.summary.txt that resulted from the example input file exa1.txt. The prefix aaa of the file name is user-specified, the size of confidence intervals equals 95% of the permutation values, and number of permutations equals 1000. Because initiation of the random number generator is based on the system's time, the results of permutation analysis change from run to run, but they should stabilize as the number of permutations is increased.

Table 2: Output file aaa.summary.txt for input file exa1.txt and prefix aaa. Continued on next pages.

```
Program    : DifferInt (May 2013) compiled by GNU Fortran (SUSE Linux) 4.3.2
Input file : exa1.txt
Begin      : 16.05.2013 at 10:32:46
-------------------------------------------------------------------------------
OBSERVED_DIFFERENTIATION_OVER_THREE_LEVELS_OF_INTEGRATION
FOR_LOCI    1   2
-------------------------------------------------------------------------------
Sample_size       = No._individuals_with_no_missing_alleles
Population_weight = 1/No.populations
Pop         1     2     3
SmplSiz    22    10    18
Weight  0.333 0.333 0.333
-------------------------------------------------------------------------------
GENE_POOL
     Complementary   Dispersive      Pairwise
     differentiation differentiation distances
Popj Delta_SD(j)     Delta-bar(j)    Delta
     =delta_SD(j)    =delta-bar(j)   =d0
                                       1       2       3
     ---------------------------------------------------
  1  0.1564          0.2095       |  0.0000  0.0705  0.3485
  2  0.2424          0.2436       |  0.0705  0.0000  0.4167
  3  0.3826          0.3826       |  0.3485  0.4167  0.0000
     ---------------------------------------------------
     Delta_SD        Delta-bar
     =delta_SD       =delta-bar
     0.2605          0.2785
-------------------------------------------------------------------------------
MEAN_SINGLE_LOCUS_GENOTYPES - CONSIDERING_ELEMENTARY_GENIC_DIFFERENCES
     Complementary   Dispersive      Pairwise
     differentiation differentiation distances
Popj Delta_SD(j)     Delta-bar(j)    Delta
                                       1       2       3
     ---------------------------------------------------
  1  0.1977          0.2455       |  0.0000  0.0932  0.3977
  2  0.2424          0.2549       |  0.0932  0.0000  0.4167
  3  0.4053          0.4072       |  0.3977  0.4167  0.0000
     ---------------------------------------------------
     Delta_SD        Delta-bar
     0.2818          0.3025
```

MEAN_SINGLE_LOCUS_GENOTYPES - NEGLECTING_ELEMENTARY_GENIC_DIFFERENCES

Popj	Complementary differentiation delta_SD(j)	Dispersive differentiation delta-bar(j)	Pairwise distances d0		
			1	2	3
1	0.2742	0.3202	0.0000	0.1227	0.5177
2	0.3232	0.3391	0.1227	0.0000	0.5556
3	0.5328	0.5366	0.5177	0.5556	0.0000

delta_SD	delta-bar
0.3768	0.3987

MULTILOCUS_GENOTYPES - CONSIDERING_ELEMENTARY_GENIC_DIFFERENCES

Popj	Complementary differentiation Delta_SD(j)	Dispersive differentiation Delta-bar(j)	Pairwise distances Delta		
			1	2	3
1	0.2396	0.2659	0.0000	0.1227	0.4091
2	0.2427	0.2711	0.1227	0.0000	0.4194
3	0.4092	0.4143	0.4091	0.4194	0.0000

Delta_SD	Delta-bar
0.2972	0.3171

MULTILOCUS_GENOTYPES - NEGLECTING_ELEMENTARY_GENIC_DIFFERENCES

Popj	Complementary differentiation delta_SD(j)	Dispersive differentiation delta-bar(j)	Pairwise distances d0		
			1	2	3
1	0.5182	0.5273	0.0000	0.2364	0.8182
2	0.4354	0.5126	0.2364	0.0000	0.7889
3	0.7934	0.8035	0.8182	0.7889	0.0000

delta_SD	delta-bar
0.5823	0.6145

```
================================================================================
                        P_E_R_M_U_T_A_T_I_O_N___A_N_A_L_Y_S_I_S
          PERMUTATION_OF_ALLELES_OVER_INDIVIDUALS__W_I_T_H_I_N__POPULATIONS
--------------------------------------------------------------------------------
1000 PERMUTATIONS_OF_ALLELES_OVER_INDIVIDUALS__W_I_T_H_I_N__POPULATIONS
FOR_ALL   2 LOCI
AT_LEVELS_OF_GENETIC_INTEGRATION  GenePool
                                  MnSLGeno=Mean_Single_Locus_Genotypes
                                  MultGeno=Multilocus_Genotypes
--------------------------------------------------------------------------------
DIFFERENTIATION                Observed   ------------ Permutation_analysis ------------
                                 values     Min  0.950-ConfidenceInterval   Max   P-Value
Considering_elementary_genic_differences:
Complementary-Differentiation-Delta_SD
          GenePool              0.2605   ................ not_affected .................
          MnSLGeno              0.2818   0.2605  [  0.2626 ,  0.3083 ]   0.3324   0.3000
          MultGeno              0.2972   0.2658  [  0.2776 ,  0.3275 ]   0.3479   0.5430
Dispersive-Differentiation-Delta-bar
          GenePool              0.2785   ................ not_affected .................
          MnSLGeno              0.3025   0.2805  [  0.2830 ,  0.3396 ]   0.3626   0.5010
          MultGeno              0.3171   0.2865  [  0.3002 ,  0.3524 ]   0.3742   0.6560
Neglecting_elementary_genic_differences:
Complementary-Differentiation-delta_SD
          GenePool              0.2605   ................ not_affected .................
          MnSLGeno              0.3768   0.3710  [  0.3932 ,  0.5177 ]   0.5729   0.9950
          MultGeno              0.5823   0.5658  [  0.6195 ,  0.8116 ]   0.9212   0.9980
Dispersive-Differentiation-delta-bar
          GenePool              0.2785   ................ not_affected .................
          MnSLGeno              0.3987   0.3867  [  0.4264 ,  0.5746 ]   0.6338   0.9970
          MultGeno              0.6145   0.6101  [  0.6653 ,  0.8539 ]   0.9394   0.9990
--------------------------------------------------------------------------------
COVARIATION_OF_DIFFERENTIATION Observed   ------------ Permutation_analysis ------------
                                 values     Min  0.950-ConfidenceInterval   Max   P-Value
Considering_elementary_genic_differences:
Cov(Delta_SD(j))
          GenePool-MnSLGeno     1.0000   0.7995  [  1.0000 ,  1.0000 ]   1.0000   0.0000
          GenePool-MultGeno     1.0000   0.8387  [  1.0000 ,  1.0000 ]   1.0000   0.0000
          MnSLGeno-MultGeno     1.0000   0.9912  [  1.0000 ,  1.0000 ]   1.0000   0.0000
Cov(Delta-bar(j))
          GenePool-MnSLGeno     1.0000   0.9479  [  0.9969 ,  1.0000 ]   1.0000   0.0000
          GenePool-MultGeno     1.0000   0.9397  [  0.9969 ,  1.0000 ]   1.0000   0.0000
          MnSLGeno-MultGeno     1.0000   0.9904  [  1.0000 ,  1.0000 ]   1.0000   0.0000
Neglecting_elementary_genic_differences:
Cov(delta_SD(j))
          GenePool-MnSLGeno     1.0000   0.7404  [  0.9890 ,  1.0000 ]   1.0000   0.0000
          GenePool-MultGeno     0.8809   0.7197  [  0.9388 ,  1.0000 ]   1.0000   0.9870
          MnSLGeno-MultGeno     0.9460   0.6162  [  0.9438 ,  1.0000 ]   1.0000   0.9740
Cov(delta-bar(j))
          GenePool-MnSLGeno     1.0000   0.9165  [  0.9947 ,  1.0000 ]   1.0000   0.0000
          GenePool-MultGeno     0.9887   0.6936  [  0.9741 ,  1.0000 ]   1.0000   0.9340
          MnSLGeno-MultGeno     0.9953   0.7205  [  0.9579 ,  1.0000 ]   1.0000   0.8720
```

```
================================================================================
                          P_E_R_M_U_T_A_T_I_O_N___A_N_A_L_Y_S_I_S
              PERMUTATION_OF_ALL_INDIVIDUAL_GENOTYPES__A_M_O_N_G__THE_POPULATIONS
--------------------------------------------------------------------------------
1000 PERMUTATIONS_OF_ALL_INDIVIDUAL_GENOTYPES__A_M_O_N_G__POPULATIONS
FOR_ALL   2 LOCI
AT_LEVELS_OF_GENETIC_INTEGRATION  GenePool
                                  MnSLGeno=Mean_Single_Locus_Genotypes
                                  MultGeno=Multilocus_Genotypes
--------------------------------------------------------------------------------
DIFFERENTIATION                 Observed   ------------ Permutation_analysis ------------
                                  values     Min  0.950-ConfidenceInterval   Max   P-Value
Considering_elementary_genic_differences:
Complementary-Differentiation-Delta_SD
          GenePool              0.2605   0.0662  [  0.0998 ,  0.2528 ]   0.3018   0.0150
          MnSLGeno              0.2818   0.1011  [  0.1383 ,  0.2798 ]   0.3239   0.0230
          MultGeno              0.2972   0.1343  [  0.1675 ,  0.2966 ]   0.3481   0.0250
Dispersive-Differentiation-Delta-bar
          GenePool              0.2785   0.0801  [  0.1137 ,  0.2862 ]   0.3391   0.0420
          MnSLGeno              0.3025   0.1198  [  0.1598 ,  0.3198 ]   0.3665   0.0510
          MultGeno              0.3171   0.1545  [  0.1942 ,  0.3378 ]   0.3774   0.0780
Neglecting_elementary_genic_differences:
Complementary-Differentiation-delta_SD
          GenePool              0.2605   0.0662  [  0.0998 ,  0.2528 ]   0.3018   0.0150
          MnSLGeno              0.3768   0.1535  [  0.2033 ,  0.3817 ]   0.4786   0.0310
          MultGeno              0.5823   0.3386  [  0.4125 ,  0.6291 ]   0.7093   0.1340
Dispersive-Differentiation-delta-bar
          GenePool              0.2785   0.0801  [  0.1137 ,  0.2862 ]   0.3391   0.0420
          MnSLGeno              0.3987   0.1815  [  0.2337 ,  0.4328 ]   0.5153   0.1050
          MultGeno              0.6145   0.3859  [  0.4714 ,  0.6956 ]   0.7828   0.3050
================================================================================
Output_files:
    aaa.summary.txt
    aaa.alleles.txt
    aaa.singlegeno.txt
    aaa.multigeno.txt
    aaa.d0matrix.txt    : pairwise distances d0 between pops at all levels
    aaa.Deltamatrix.txt : pairwise distances Delta between pops at all levels
To_plot_snails,_enter_the_following_commands_in_a_terminal_window:
   xfig -allownegcoords aaa.bigdeltasd.schnecken.fig
   xfig -allownegcoords aaa.deltasd.schnecken.fig
   xfig -allownegcoords aaa.bigdeltabar.schnecken.fig
   xfig -allownegcoords aaa.deltabar.schnecken.fig
End: 16.05.2013 at 10:32:58
--------------------------------------------------------------------------------
```