GSED Version 3.0

"Genetic Structures from Electrophoresis Data"

User's Manual

April 2010

http://www.uni-goettingen.de/de/95607.html

Elizabeth M. Gillet

Abteilung Forstgenetik und Forstpflanzenzüchtung

Universität Göttingen

Büsgenweg 2 / 37077 Göttingen / Germany

Email egillet@gwdg.de

GSED Version 3.0 User's Manual, April 2010

Revision of GSED User's Manual of April 1998 ©E.M. Gillet

Download at URL http://www.uni-goettingen.de/de/95607.html

Author:
Elizabeth M. Gillet
Abt. Forstgenetik und Forstpflanzenzüchtung
Universität Göttingen
Büsgenweg 2
37077 Göttingen, Germany
Email: egillet@gwdg.de

Contents

List of Tables

1 Introduction

The purpose of GSED ("Genetic Structures from Electrophoresis Data") is to characterize the genetic variation observed in one or more demes of individuals of the same species (e.g. populations, stands, ontogenetic stages, generations). Alleles can be coded by any non-negative integers, thus allowing for the designation of microsatellite alleles by their numbers of base-pairs. For any combination of gene loci and any of the genetic structures (*i.e.*, frequency distributions) that can be constructed from the alleles, haplotypes, gene pools, or genotypes at any combination of loci in the demes (see Table 1), GSED calculates measures of genetic variation (see Table 2). These variation measures are based on a conceptually and mathematically unified system of data analysis for population genetic investigations that has been and continues to be developed at the Institut für Forstgenetik und Forstpflanzenzüchtung of the Universität Göttingen and at the Institut für Populations- und ökologische Genetik (http://www.ipoeg.de) in Göttingen.

Table 1: *Genetic structures calculated by GSED*

Single locus	Allele frequencies among maternal contributions*
	Allele frequencies among paternal contributions**
	Allele frequencies***
	Genotype frequencies
Multilocus	Haplotype frequencies among maternal contributions*
	Haplotype frequencies among paternal contributions**
	Haplotype frequencies***
	Genotype frequencies

* if gametic sex is inferable e.g. as the allele or multilocus haplotype contributed by the maternal/seed parent to the megagametophyte of conifer seeds, to chloroplasts, or to mitochondria (angiosperms)
** if gametic sex is inferable e.g. as the allele or multilocus haplotype contributed by the paternal/pollen parent to the embryo of conifer seeds as determined by megagametophyte/embryo analysis or to chloroplasts (conifers)
*** The combined set of all maternal and paternal alleles or multilocus haplotypes

1.1 Genetic structures and their characterization

The foundation of this system of data analysis is the quantification of differences between demes as the proportion of individuals that must be changed in one of the demes to make its genetic structure match the structure in the other deme. This descriptive concept of difference is applicable to all types of demes in any situation, since it does not rely on assumptions of specific models (e.g. drift, lack of mutation, special mating systems).

Table 2: *Characterization of genetic structures by GSED*

- Analysis of allelic, haplotypic and genotypic structures
  - *Measures of variation within demes*
    - Diversity v_2
    - Total population differentiation δ_T
    - Evenness e
  - *Measures of variation between demes*
    - Genetic distance d_0
    - Subpopulation differentiation D_j and δ
    - Test of homogeneity
- Analysis of genotypic structure
  - Heterozygosity, single locus and multilocus, observed and conditional
  - Test of Hardy-Weinberg structure and heterozygosity
  - Test of product structure
- Analysis of the gene pool
  - *Measures of variation within demes*
    - Diversity v_2 of the gene pool
    - Diversity v_{gam} of the hypothetical gametic output
    - Total population differentiation δ_T of the gene pool
  - *Measures of variation between demes*
    - Distance d_0 between gene pools
    - Differentiation δ of subdivided gene pools

measure of absolute distance

$$d_0(P, P') = \frac{1}{2} \sum_{i=1}^{n} \mid p_i - p'_i \mid$$

between two demes (e.g. populations) P and P', where p_i and p'_i denote the relative frequency of individuals of type i in deme P and P', respectively, with respect to a trait that is expressed in each individual as one of n types (trait states) (Gregorius, 1974a,b, 1978, 1984a). d_0 ranges from $d_0 = 0$ for demes with identical frequency distributions to $d_0 = 1$ for disjoint demes, *i.e.*, demes that share no types. The metric distance d_0 quantifies the proportion of individuals in one of the demes whose type must be changed in order to make this deme match the other. d_0 can be applied to genetic traits and to phenotypic traits whose genotypes have yet to be determined.

For genetic traits, the p_i refer to genetic types that can be of arbitrary complexity. Thus d_0 enables comparison of demes at any level of genetic integration of the underlying genes: from the lowest level of the alleles at a single locus – to the level of the gene pool over loci – to the level of multilocus genotypes – to the level of multilocus haplotypes. Inference of multilocus haplotypes requires specification of the gametic sex of each allele, for example, as the cytotypes of uniparentally inherited (haploid) organelles or as the multilocus (haploid) gametic contribution of the maternal (seed) parent or the paternal (pollen) parent of conifer seeds, where the gametic sex of each allele is determinable by means of megagametophyte/embryo analysis. GSED calculates genetic structures as the frequency distributions of the genetic types at any chosen level of integration from lists of multilocus genotypes (multilocus haplotypes require specification of the gametic sex of the alleles).

For each genetic structure, GSED calculates measures of variation within and between demes, most of which are based on d_0. For example, matrices of pairwise distances d_0 are calculated that can be imported into programs that construct dendrograms. For demes that can be considered as subpopulations of a large population, d_0 forms the basis of the measure δ of subpopulation differentiation. δ measures the mean genetic distance D_j of each subpopulation j to its complement that is formed by pooling all other subpopulations (Gregorius & Roberds, 1986). As a true measure of differentiation, $\delta = 0$ when all subpopulations are identical and $\delta = 1$ when all subpopulations are genetically disjoint, *i.e.*, share no types.

d_0 is useful to quantify variation not only between demes but also within demes. In the special case in which each individual is considered to form a subpopulation of its own, δ reduces to the measure of total population differentiation δ_T (Gregorius, 1987, 1988). Measures of evenness specify the minimum genetic distance to a uniform distribution (Gregorius, 1990). Only the diversity measure v_2 does not rely on d_0.

The variation within individual genotypes in a deme at singe loci is measured as the proportion of heterozygous individuals, both observed and conditional on the allele frequencies. For multiple loci, the distribution of the number of heterozygous loci per individual is given.

In addition to the calculation of genetic structures and variation measures that are not to be found in other software, GSED includes several statistical tests: (1) A test of homogeneity

among demes at any level of genetic integration provides an additional measure of between-deme variation. (2) Tests of the correspondence of the manner in which alleles are associated in genotypes to special mating systems: (a) A test of Hardy-Weinberg proportions examines the hypothesis of random mating within a deme. (b) A test of product structure examines the hypothesis of random fusion between asymmetric gametic distributions in cases where the gametic sex of the alleles is known.

Current research in the field of population genetics underlines the importance of this "alternative" system of data analysis. The realization is spreading that the most commonly used measure F_{ST} $(= G_{ST})$ (Wright, 1978) does not measure differentiation among populations, when differentiation is understood in the sense of differences (Gregorius, 1987; Jost, 2008). A new conceptual analysis of the distribution of variation over populations shows that whereas F_{ST} is not a measure of differentiation among populations but rather a measure of the apportionment of variation to populations, δ is indeed a measure of differentiation.

Many of the measures of variation calculated by GSED can be applied not only to genetic types but also to any system of classification by which each individual of a population can be assigned one of a finite set of discrete "types" (e.g., phenotypes, ecotypes). Although the assumption that data input to GSED concerns genetic types is reflected in its commentaries, one or higher dimensional non-genetic classifications can be disguised as maternal alleles or haplotypes at "loci" for which gametic sex is specified and "paternal" type unknown. An input file would be analogous to that of Example 4 in App. C. Output headings would have to be reinterpreted accordingly.

1.2 Requirements on the data

Input to GSED consists of a list of genotypes or haplotypes scored in individuals belonging to one or more collections, or demes. An individual's **haplotype** or **genotype** refers to a single allele or pair of alleles, respectively, that is/are present at a single gene locus (**single locus haplotype/genotype**) or at each of two or more gene loci (**multilocus haplotype/genotype**). In diploid organisms, the **gametic sex** of an allele at a locus can sometimes be determined as the contribution of the female or the male gametophyte to the nuclear or organelle genome (see legend of Table 1). If the **gametic sex** of each allele at each locus, that is, the sex of the contributing parent, is specifiable, the genotype can be designated as an **ordered genotype**, with the maternal allele in the first position and the paternal in the second.

When using GSED to analyze genetic types, it is essential that the alleles at each locus be known. In other words, the phenotype produced by the genes at each locus must be a **gene marker**, in that the phenotype enables identification of *all* involved alleles. Microsatellite and isoenzyme phenotypes that result from gene loci showing a codominant mode of inheritance are gene markers. For a dominant mode of inheritance, such as is caused by the presence of a (recessive) null allele at a locus, the phenotypes do not define a gene marker; loci showing dominance cannot be used for the analysis of genetic types unless additional inheritance analysis has revealed the true genotype of each individual

at the locus. Data that is simply missing (denoted as the "allele -1") in an individual's genotype causes the individual to be ignored in calculations.

1.3 Implementation

GSED reads input data from a text file that was prepared using an editor or spreadsheet. The user interactively chooses which frequency distributions are determined from the data and which calculations are performed. In this newest version of GSED, two interactive modes are provided: (1) Menu-directed[1] choice of input file, genetic structures, measures, and tests; (2) alternatively, those who prefer keyboard entry may choose to answer queries as in earlier versions of GSED. Two text files are produced as output, one of which contains the complete output. The other lists the measures of variation in a compact form that is designed to be imported into a spreadsheet program (e.g. Excel, OpenOffice).

GSED is written in FORTRAN and compiled using a GNU Fortran95 compiler. Executables for the operating systems Win and Linux (openSUSE 11.1) are available for downloading at http://www.uni-goettingen.de/de/95607.html. Version 3.0 of GSED contains several major improvements over earlier versions: (1) Menu-directed choice of options; (2) simplification of the format of the input data file and the allowance for commentary lines; (3) improvement of the importability of the tabular output file into spreadsheet programs.

1.4 Organization of this manual

The following sections (2, 3, 4) of this manual deal with practical matters, namely construction of an input file, execution of the program, and understanding the output. The following four sections (5, 6, 7, 8) are concerned with the concepts behind the program. The first of these sections (5) reviews the different types of frequency distribution that can be calculated from demes of multilocus genotypes. Three sections (6, 7, 8) outline the measures and tests that are performed for the various frequency distributions, including references to (mostly original) articles containing detailed descriptions of the underlying concepts. A list of references follows, denoted in the text by numbers in square brackets. Appendices contain technical specifications (compiler, limitations on data) and examples of input and output files.

[1] using graphical user interface (GUI) routines from the scientific data plot software DISLIN (Michels, 2009)

2 Constructing an input file

A GSED input file is constructed using any text editor, word processor, or spreadsheet program. The input file must contain only ASCII characters, which means that any formatting information must be eliminated before running GSED:

- Simple text editors automatically save the data in the correct form as a text-only file with default extension `.txt`.
- For word processors, formatting information is eliminated by saving the input as a text file, usually with extension `.txt`.
- For spreadsheets, each line of input can be constructed by putting each piece of data (name or number) into a field of its own. After saving the data in the normal spreadsheet format to enable later revision, the GSED input file is made by saving the input as a CSV text file with the extension `.csv` or, if available, directly as a text-only file with extension `.txt`. The field separator can be chosen to be a blank, a comma, or a tab-character.

After saving, changing the extension of an input file to `.dat` or `.inp`, for example, may help to distinguish it from GSED output files, which have the extension `.txt`.

Each input file consists of three parts:

- The **header** that defines the numbers and names of the gene loci and demes (see 2.1);
- The **READ format line** that specifies how the lines of data are to be interpreted (see 2.2);
- The **deme data** containing the genotypes in each deme.

Examples of input files are given in App. C.

2.1 Header

The header, which occupies the first lines of an input file, defines the number and names of demes and gene loci. It consists of:

- One line containing the number of demes and the number of gene loci, separated by a blank, comma, or tab-character;
- One line per gene locus, containing the name of the locus ($\leq$ 12 characters). The first-named locus is referred to as Locus 1 in the remainder of the data, the second-named locus as Locus 2, etc.
- One line per deme, containing the name of the deme ($\leq$ 40 characters) The first-named deme is referred to as Deme 1 in the remainder of the data, the second-named deme as Deme 2, etc.

2.2 READ format line

All data in the deme specification line and the genotype lines are integers. In particular, the allele designations, or "names", must be **non-negative integers**. No non-numeric letters are permitted. The allele designation "0" is meant to denote a "null allele", if its presence can be determined.

The READ format line of an input file specifies how the integers in the deme specification line and the genotype lines are to be read (see 2.3). It can have one of two forms:

2.2.1 Unformatted input

If each integer in the genotype lines is followed by a field separator (blank, comma, or tab-character), as in CSV-files, the READ format line can simply be specified by placing the word "`unformatted`" anywhere within the first 70 columns of the line.

Advantages and disadvantages of unformatted input:

- Advantage: The integers can appear anywhere on the data line, as long as they are in the correct order and separated by a blank, comma, or tab-character.
- Disadvantage: No additional information can be included on the data lines, such as non-numeric text.

This line can be followed by one or more empty lines.

2.2.2 Formatted input

The **READ format line** for formatted input specifies in which columns each integer of the data is to be found. It has the following form:

- Columns 1-2, right-justified: The number of loci specified on each line of data
- Beginning in column 5: A FORTRAN READ format defining which columns in each of the subsequent data lines contain each of $2 + 2 * n$ integers, where n is the number of gene loci (see 2.2).

Advantages and disadvantages of formatted input:

- Advantages:
  - Shorter data lines, since separators (blanks, commas, tab-characters) between integers are not needed.
  - Inclusion of non-numeric text in specified columns of the data lines, since the FORTRAN READ format can be constructed to skip over these columns.

- Disadvantage:
  - The integers must appear in the same columns in every data line, as specified by the FORTRAN READ format.

Examples of FORTRAN READ formats are shown in 2.3. The following box gives a short description of each of the elements of a FORTRAN READ format that are of relevance here (for further information consult any FORTRAN language reference manual):

`nIw`	The `I` field descriptor indicates that an integer is to be read in a field of width `w` (*i.e.*, `w` columns). The `n` specifies the number of integers that are to be read in consecutive fields of width `w`.
`nX`	The `X` field descriptor indicates that `n` columns are to be skipped
`r(...)`	The repeat count `r` indicates that the contents of the parentheses are to be repeated `r` times
`,`	Separates field descriptors
`/`	Separates field descriptors and causes reading to continue on a new line
`(...)`	Parentheses enclose the entire FORTRAN READ format

Examples of FORTRAN READ formats: *Example 1.*

```
10  (2I4,1X,10(2I2))
.........+.........+
```

reads 22 integers, including a 10-locus genotype, from the following line of data (namely 2, 123, 1, 2, 2, 3, .., 3, 3):

```
   2 123  1 2  2 3  3 5  2 1  2 3  2 1  1 1  1 1  2 2  3 3
.........+........+.........+.........+.........+.........+
```

Example 2.

```
20  (I4,I4,10(2I2)/9X,10(2I2))
.........+.........+.........+
```

reads 42 integers, including a 20-locus genotype, from the following two lines of data (2, 2345, 2, 1, 2, 3, .., 1, 1):

```
   22345  21 23 24 22 33 11 21 21 23 21
          32 33 12 23 23 31 32 23 32 11
.........+........+.........+.........+
```

2.3 Deme data

The third part of an input file consists of the deme data, *i.e.*, a block of data lines for each of the demes.

Allele designations, or "names", must be **non-negative integers**. The allele "0" is meant to denote a "null allele", if its presence can be determined. A missing allele is denoted by "`-1`". Alleles can thus be designated as "`1, 2, 3, ...`", as is common for isoenzymes, or for example as the numbers of base-pairs "`101, 122, 143, ...`" of microsatellite alleles.

Deme data specifies the single-locus or multilocus **genotypes** or **haplotypes** in each deme (see Tab. 3). Specification of (diploid) genotypes requires knowledge of both alleles at the locus, *i.e.*, which in general necessitates the codominance of the mode of inheritance (see Gillet (1996)).

If the **gametic sex** of the alleles making up a single-locus genotype is known (such as for the combined megagametophyte/embryo analysis of conifer seeds), the genotypes can be designated as **ordered genotypes**; in this case, the maternal allele is listed first and the paternal allele second. Organelle **cytotypes** can be specified as ordered genotypes in which the allele of one gametic parent is unknown.

If more than one locus is scored in the same individual, its **multilocus genotype** is specified by a line of data listing its single-locus genotypes. If gametic sex is specified at all of the loci, the multilocus **maternal haplotypes** and **paternal haplotypes** can be inferred. **Multilocus organelle cytotypes** are represented as ordered multilocus genotypes with one parental allele unknown at all loci.

The data for each deme consists of three parts, each of which is explained in detail in the following subsections:

- A **deme specification line** defining the order and the gametic sex specification of the loci
- Lines containing each of the **genotypes** found in the deme and, if indicated in the deme specification line, the number of individuals possessing this genotype
- The **end-of-deme** line

Demes can appear in any order, since the deme number is included on each line of data.

2.3.1 Deme specification line

The data for each deme begins with the **deme specification line**. This line contains the following information (as integers) about each deme:

- The deme number, in accordance with the ordering of the deme names in the header

Table 3: *Genetic types that can be represented in GSED input data:*

- **Single-locus genotype**: The pair of alleles at a single gene locus
- **Ordered single-locus genotype**: Single-locus genotype in which the allele inherited from the maternal parent (**maternal allele**) is listed first and the allele inherited from the paternal parent (**paternal allele**) second; requires specification of the **gametic sex** of the alleles
- **Single-locus maternal haplotype**: The allele contributed by the maternal parent, listed in the first position of an ordered genotype
- **Single-locus paternal haplotype**: The allele contributed by the paternal parent, listed in the second position of an ordered genotype
- **Single-locus cytotype**: The genetic variant at an organelle locus, represented as one of the alleles of an ordered single-locus genotype, the other allele being unknown
- **Multilocus genotype**: List specifying the single-locus genotype at each of a given set of loci
- **Ordered multilocus genotype**: List specifying the ordered single-locus genotypes at each of a given set of loci
- **Multilocus maternal haplotype**: List specifying the maternally-contributed allele at each of the loci in an ordered multilocus genotype, given in the first position of every single-locus genotype
- **Multilocus paternal haplotype**: List specifying the paternally-contributed allele at each of the loci in an ordered multilocus genotype, given in the second position of every single-locus genotype
- **Multilocus organelle cytotype**: List specifying the genetic variant at each of the loci in an ordered multilocus genotype, given in the first position of every single-locus genotype, the allele at the second position being unknown (or *vice versa*)

- The assignment of fields (or blocks of columns in a line of data) to the alleles at gene loci
- Specification of whether each genotype refers to a single individual (see Sec. 2.3.2) or to a given number of individuals (see Sec. 2.3.3)
- Indication of whether or not the gametic sex of the alleles at each of the loci is specified (see 5)

The deme specification line specifies the positions on the data line of the following integers for n gene loci:

$$\mathit{DemeNo} \quad \texttt{0} \quad \pm\mathit{LocusNo}_1 \quad \mathit{GamSex}_1 \quad \mathit{LocusNo}_2 \quad \mathit{GamSex}_2 \quad \ldots \quad \mathit{LocusNo}_n \quad \mathit{GamSex}_n$$

where

DemeNo	Deme number, referring to the list of demes in header
`0`	Indication that this line is a deme specification line
$\mathit{LocusNo}_1$	Number of the first locus in the genotype, referring to list of gene loci in the header. If $\mathit{LocusNo}_1 > 0$, each genotype refers to a single individual (see 2.3.2). If $\mathit{LocusNo}_1 < 0$, the second field on each genotype line gives the number of individuals that possess this genotype (see 2.3.3).
GamSex_1	Gametic sex specification of the first locus: GamSex_1="`1`" if gametic sex is specified, "`0`" otherwise
$\mathit{LocusNo}_i$ $(i = 2, \ldots, n)$	Number of the ith locus in the genotype, referring to the list of loci in the header ($\mathit{LocusNo}_i > 0$),
GamSex_i $(i = 2, \ldots, n)$	Gametic sex specification of ith locus: GamSex_i="`1`" if gametic sex is specified, "`0`" otherwise

Examples of deme specification lines for formatted input using the FORTRAN READ format

```
10   (2I4,1X,10(2I2))
.........+.........+.
```

Example 1. Deme specification line for deme 2 specifying that multilocus genotypes comprise loci 1-10, that the second integer in the subsequent genotype lines contains the "name" of the single individual whose genotype is listed, and that gametic sex is not specified for any locus:

```
   2   0  1 0  2 0  3 0  4 0  5 0  6 0  7 0  8 0  9 0 10 0
.........+.........+.........+.........+.........+.........+
```

Example 2. Deme specification line for deme 5 specifying that multilocus genotypes comprise loci 1-10, that the second integer in the subsequent genotype lines denotes the frequency of the respective genotype in the deme, and that gametic sex is specified for all loci:

```
   5   0 -1 1  2 1  3 1  4 1  5 1  6 1  7 1  8 1  9 1 10 1
.........+........+.........+.........+.........+.........+
```

2.3.2 Genotypes of single individuals

If the sign of $LocusNo_1$ in the deme specification line of a deme equals "`+`" or is blank (see 2.3.1), then each genotype is interpreted to be that of a single individual. In this case, each genotype line specifies the following integers:

DemeNo *IndivNo* $Locus_1 Allele_1$ $Locus_1 Allele_2$ ... $Locus_n Allele_1$ $Locus_n Allele_2$

where

DemeNo	Deme number, referring to the list of demes in header
IndivNo	Number designating the individual whose genotype is listed
$Locus_i Allele_1$	The first allele at locus i as an integer $\geq$"`-1`": If gametic sex is specified for this locus, then $Locus_i Allele_1$ is the allele contributed by the maternal parent.
$Locus_i Allele_2$ $(i = 2, \ldots, n)$	The second allele at locus i as an integer $\geq$"`-1`": If gametic sex is specified for this locus, then $Locus_i Allele_2$ is the allele contributed by the paternal parent

2.3.3 Genotype frequencies in deme

If the sign of $LocusNo_1$ in a deme specification line equals "`-`" (see 2.3.1), then each genotype is interpreted as having been found in a number of individuals. The frequency of a genotype in the deme is given in the second field. In this case, each genotype line specifies the following integers:

DemeNo *Frequency* $Locus_1 Allele_1$ $Locus_1 Allele_2$ ... $Locus_n Allele_1$ $Locus_n Allele_2$

where

DemeNo	Deme number, referring to the list of demes in header
Frequency	Number of individuals possessing the genotype
$Locus_i Allele_1$ $(i = 2, \ldots, n)$	First allele at locus i as an integer $\geq$"`-1`": If gametic sex is specified for this locus, then $Locus_i Allele_1$ stems from the **maternal parent**
$Locus_i Allele_2$ $(i = 2, \ldots, n)$	Second allele at locus i as an integer $\geq$"`-1`": If gametic sex is specified for this locus, then $Locus_i Allele_2$ stems from the **paternal parent**

A **null allele** at locus i is designated by $Locus_i Allele_j$="0" (zero, $j = 1$ or 2).

An **unknown allele** is specified by $Locus_i Allele_j$="-1" ($j = 1$ or 2). Note that unknown alleles in a random deme of genotypes can present a problem for the calculation of frequency distributions (see 5).

2.3.4 End-of-deme line

- For unformatted input: either an empty line or the integer "9999" somewhere on the line
- For formatted input: either an empty line or a line containing the integer "9999" in columns w-3 to w, where w is the width (Iw) of the first field defined by the FORTRAN READ format (see 2.2.2).

An end-of-deme line can (but need not) be followed by any number of empty lines before beginning the data for the next deme.

2.4 End of input

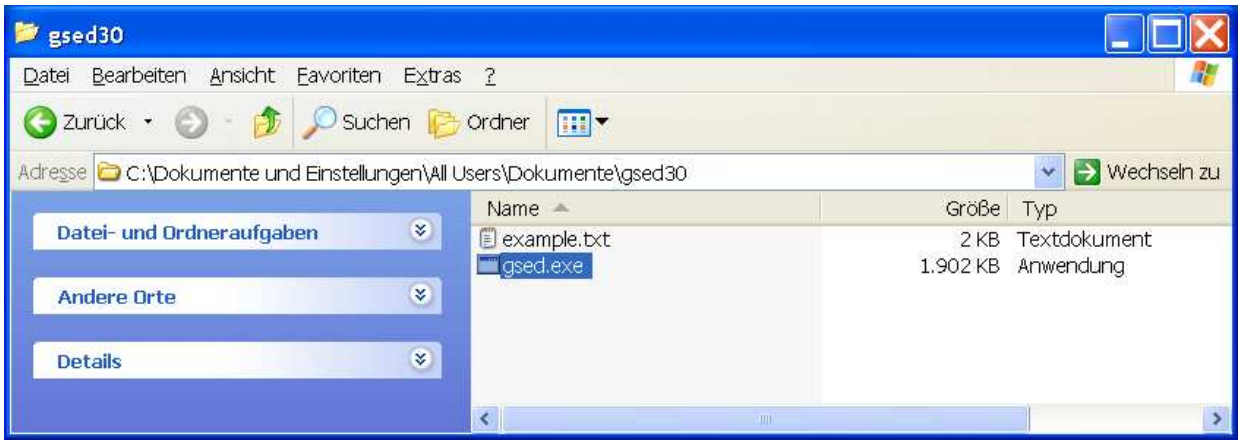
Reading of the input file terminates when no further data (*i.e.*, non-empty lines) follows an end-of-deme line.

3 Running GSED

3.1 Menu-driven execution

Version 3.0 of GSED introduces the option of menu-driven execution, thanks to routines provided by the scientific data plotting software DISLIN of H. Michels (http://www.dislin.de). To date, GSED has been compiled for WinXP and openSUSE (http://www.uni-goettingen.de/de/95607.html).

In Windows, menu-driven execution is started either by clicking on the file name `gsed.exe` in a file manager



or by opening a console window, changing the current directory to the one containing the file `gsed.exe`, and entering at the cursor

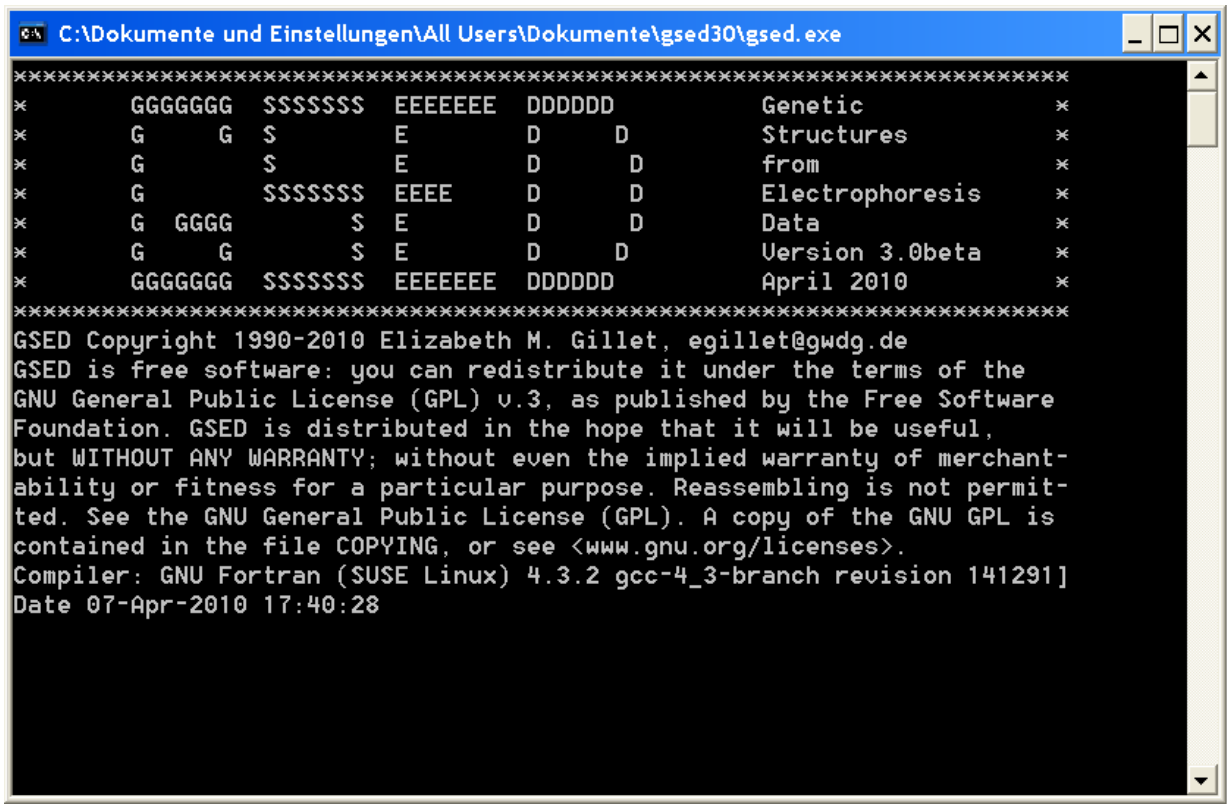
```
>gsed <Return>
```

In Linux, either click on the file `gsed.exe` in a file manager or start execution of GSED by opening a console window, changing the current directory to the one containing the file `gsed`, and entering at the cursor

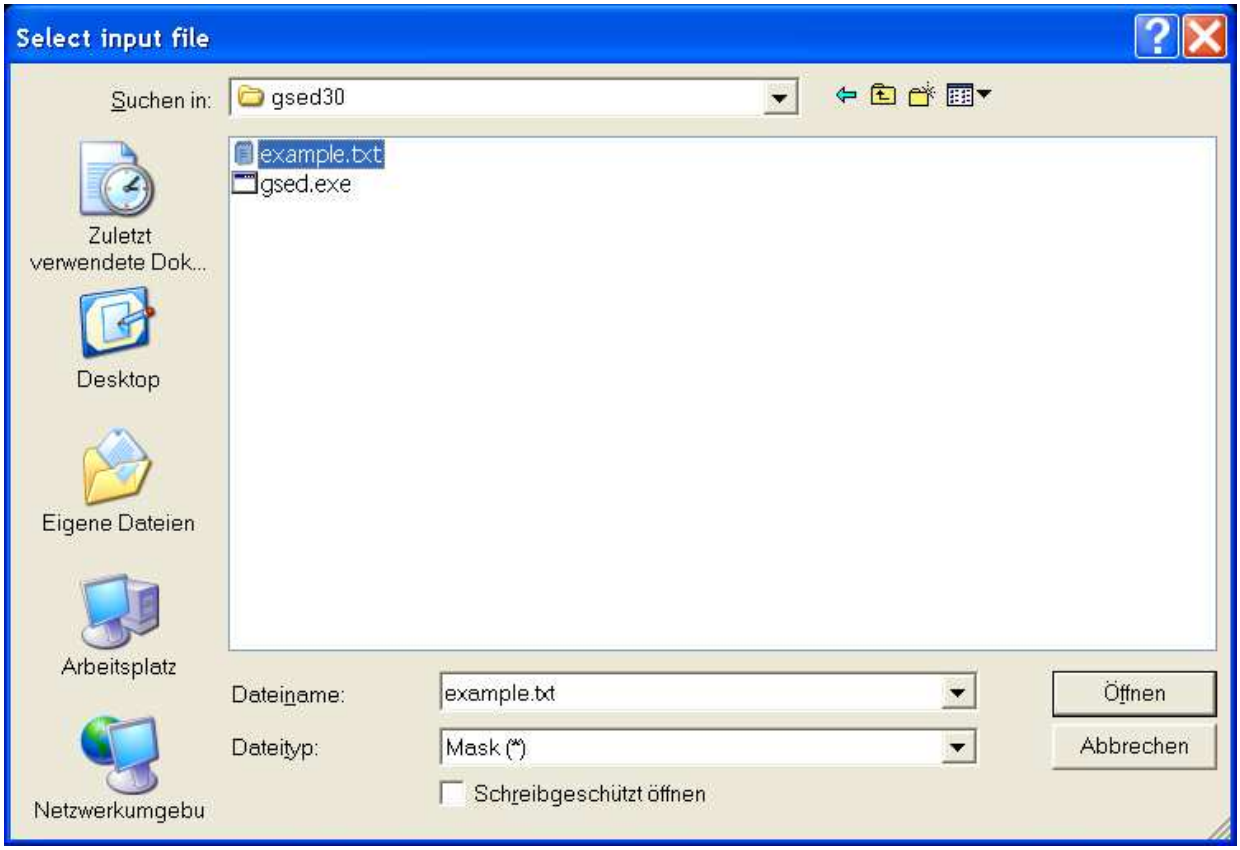
```
>./gsed <Return>
```

The following graphics illustrate execution in WinXP. In Linux, similar windows appear.

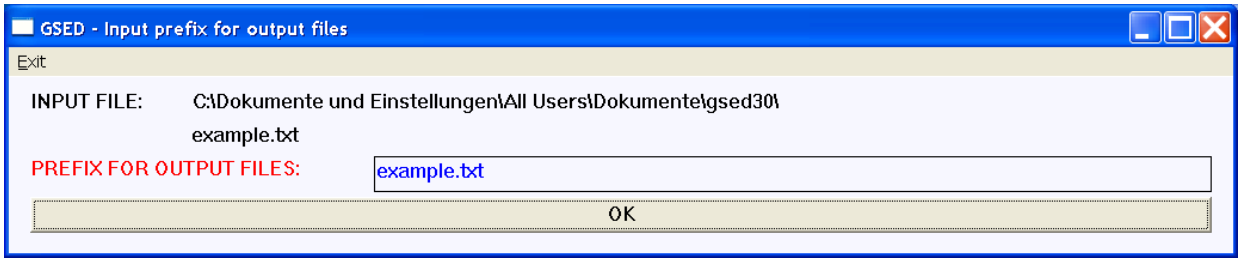
First, a console window opens that shows the GSED header, including the version numbers of GSED and the FORTRAN compiler:



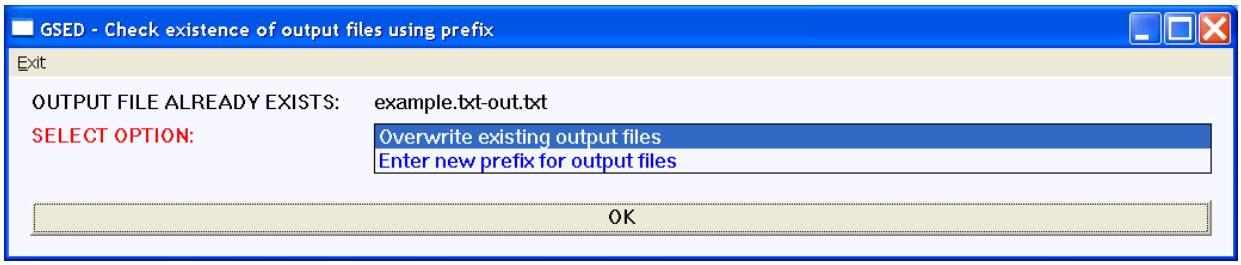
Then a File Select window opens, requesting the name of the input file. The input file need not be in the same directory as the executable `gsed.exe` (Win) or `gsed` (Linux). For illustration purposes, we clicked on the input file `example.txt` that is included in the download package.



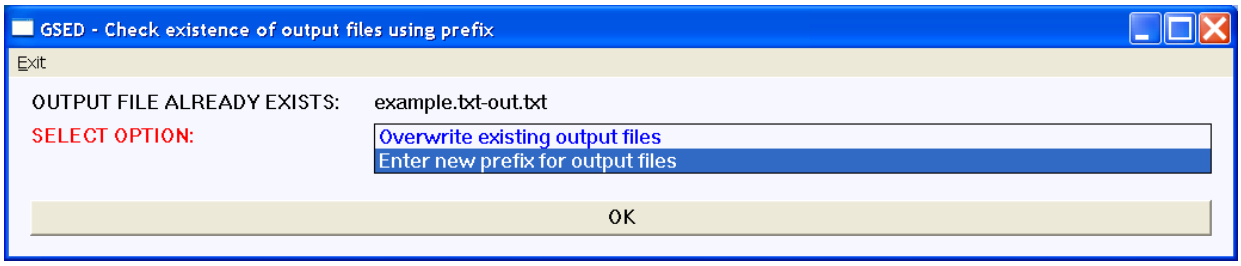
Then a window entitled "GSED - Prefix for output files" opens. It suggests that the name of the input file be used as a prefix for the names of all output files. If desired, a different prefix can be typed in. The prefix is submitted by clicking on the "OK" button.



If an output file named with the chosen prefix followed by "`-out.txt`" already exists, a window entitled "GSED - Check existence of output files using prefix" opens.

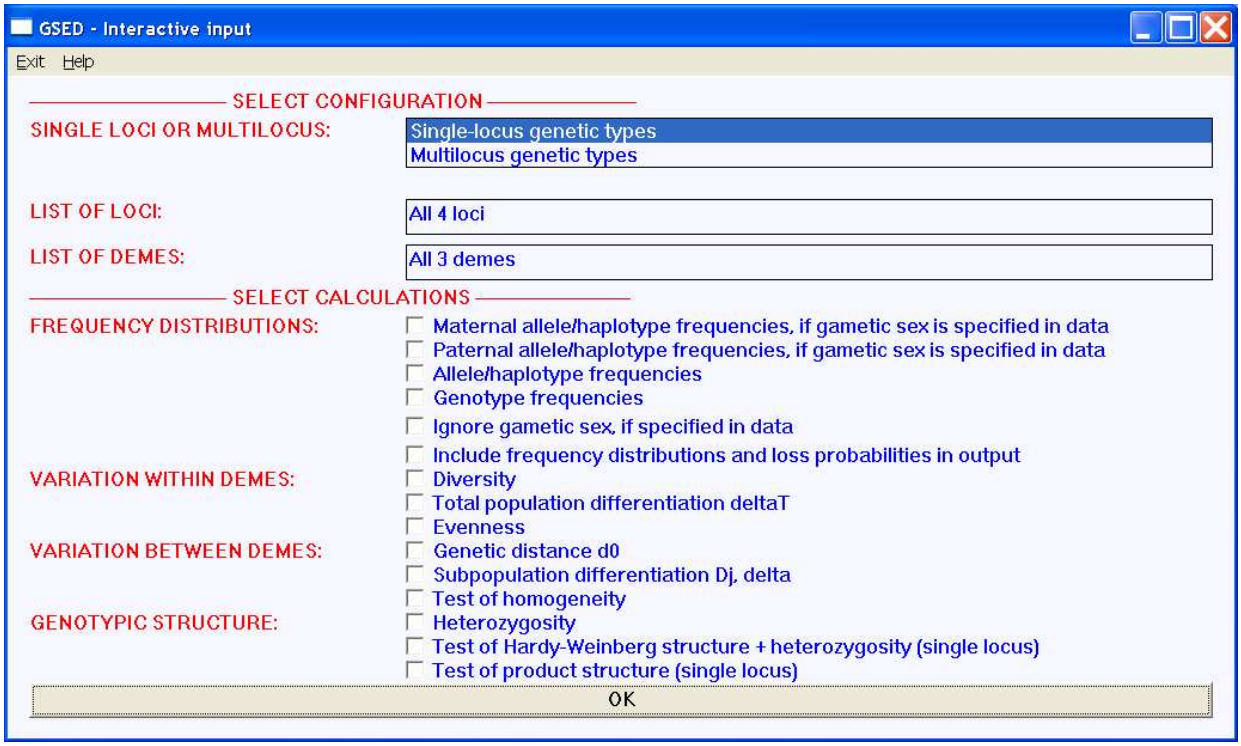


Select one of the options "Overwrite existing output files" or "Enter new prefix for output files". Again, the option is submitted by clicking on the "OK" button. If the latter option is chosen, as in this window



the window entitled "GSED - Prefix for output files" is reopened and a new prefix can be entered, such as "`xxx`". If output files with this new prefix also exist, the window entitled "GSED - Check existence of output files using prefix" reappears. This loop continues until either a prefix is entered that has not been used before or until the option "Overwrite existing output files" is selected in the "GSED - Check existence of output files using prefix" window.

The configuration and calculations are selected in the next window entitled "GSED – Interactive input".



Part 1: "— SELECT CONFIGURATION —":

In the option "SINGLE LOCI OR MULTILOCUS", the preset choice of "Single-locus genetic types" results in calculation of parameters for each single locus and for the gene pool defined by all of these loci. Choice of "Multilocus genetic types" causes all parameters to be calculated for genetic types defined by their genes at multiple loci. For example, the two-locus genetic types $A_1A_1B_1B_1$ and $A_1A_1B_1B_2$ are considered as being completely different, even though they share 3 of the 4 genes.

The next option "LIST OF LOCI" is preset to "All n loci", where n is the number of loci in the input file. If not all loci are to be chosen, change the preset text to a list of the desired loci, separated by blanks or commas, e.g. "2,3,4" or "2 3 4" (see next window).

The next option "LIST OF DEMES" is preset to "All m demes", where m is the number of demes in the input file. If not all demes are to be chosen, change the preset text to a list of the desired demes, separated by blanks or commas, e.g. "1,2" or "1 2" (see next window). A single deme can also be chosen.

Part 2: "— SELECT CALCULATIONS —":

When GSED is run with a new input file, all checkboxes are empty. For subsequent runs with an input file of the same name, the calculations that were checked in the previous run are shown (see 3.3).

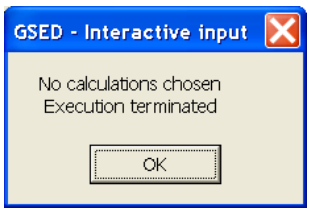
Under "FREQUENCY DISTRIBUTIONS", one or more of the distributions described in in 5 can be chosen by clicking on the checkboxes. Checking "Ignore gametic sex, if specified in data" causes GSED to treat the genotypes at the locus as unordered genotypes and to suppress calculation of maternal/paternal allele/haplotype frequencies.

Under "VARIATION WITHIN DEMES", "VARIATION BETWEEN DEMES", and "GENOTYPIC STRUCTURE", parameters are chosen for calculation (see 6).

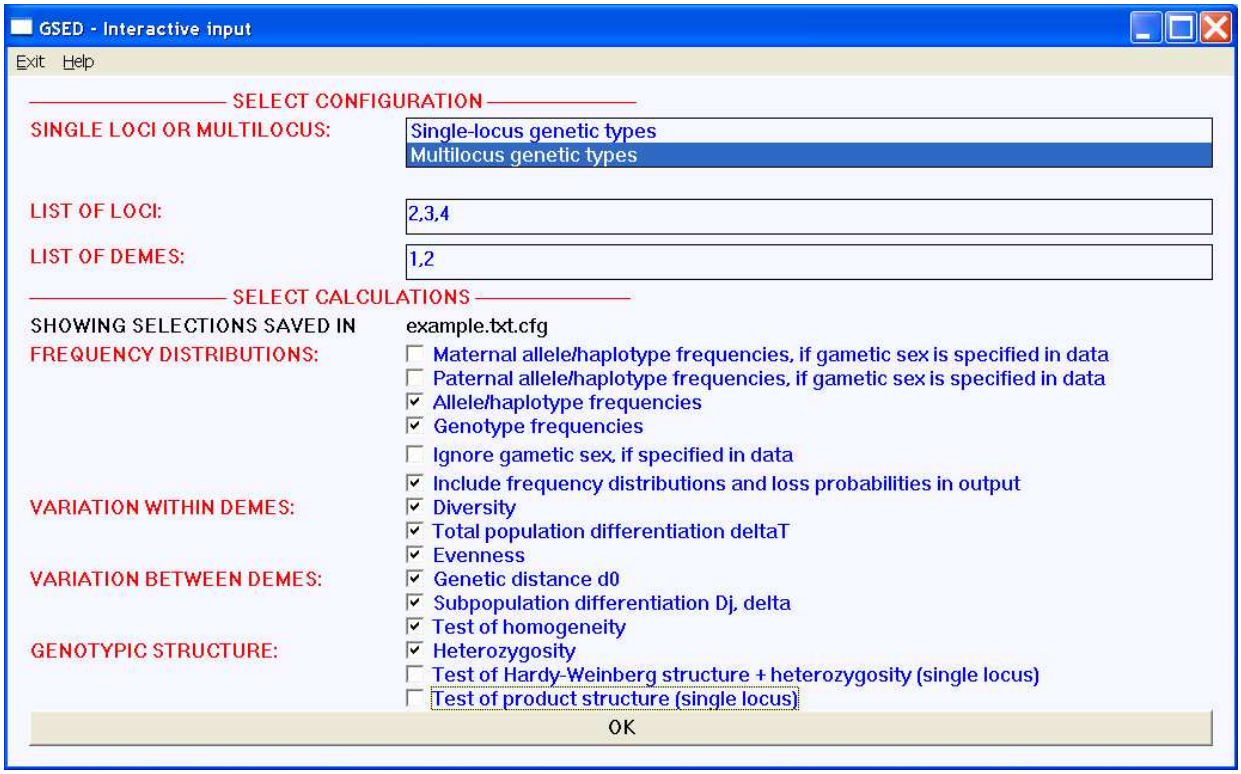
The selections are submitted by clicking on the "OK" button.

The selected calculations are saved in a configuration file, the name of which consists of the name of the input file followed by ".cfg", here "`example.txt.cfg`". In all subsequent runs using the same input file, the checkboxes in the "GSED – Interactive input" window will be preset according to the settings stored in the configuration file.

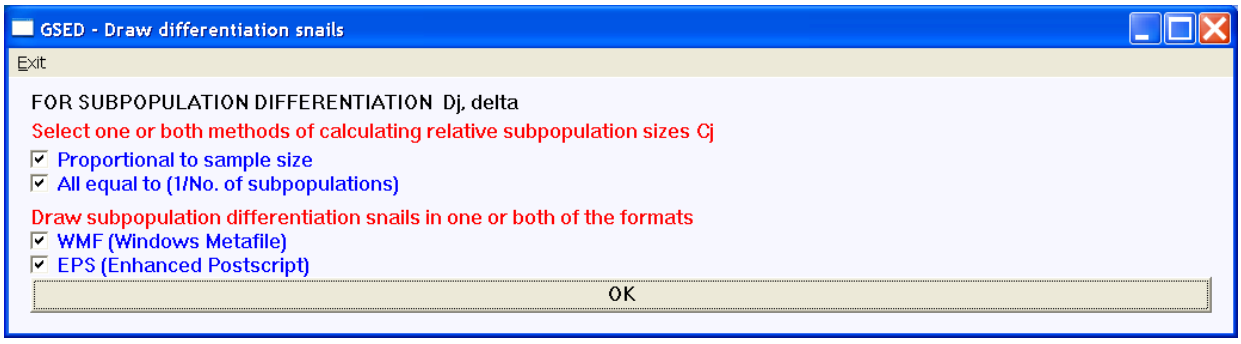
If no calculations were selected in the "GSED – Interactive input" window, the following message window appears. GSED must be restarted.



The following example shows a "GSED – Interactive input" window in which multilocus genotypes comprising loci 2, 3, and 4 are to be constructed for demes 1 and 2.

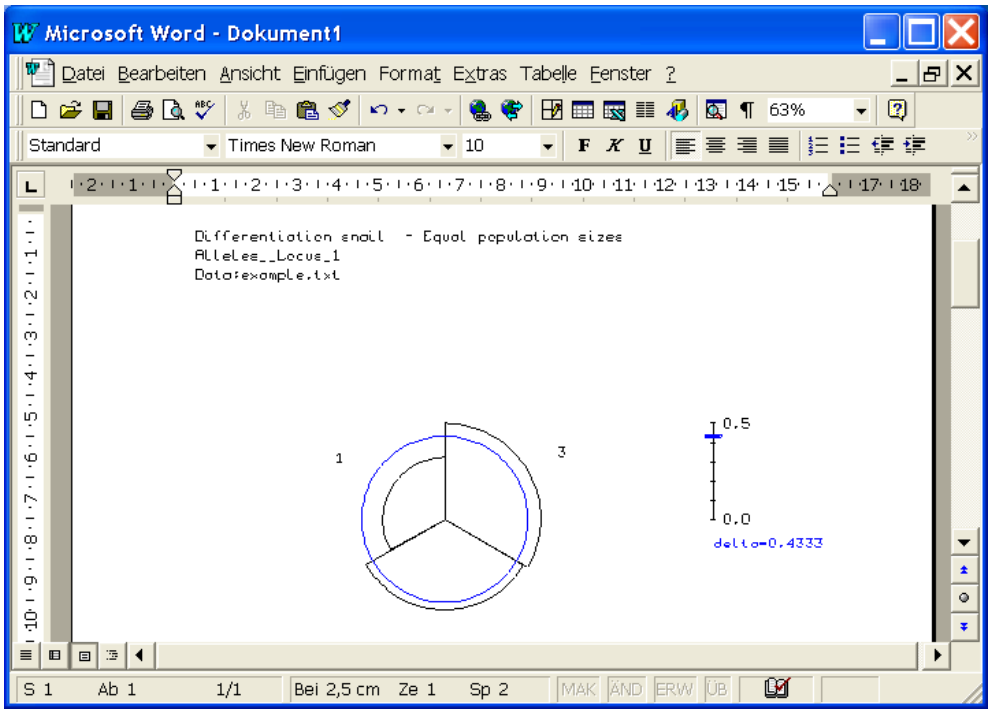


If the checkbox "Subpopulation differentiation Dj, delta" in the "GSED – Interactive input" window is marked, a window entitled "GSED – Draw differentiation snails" appears.

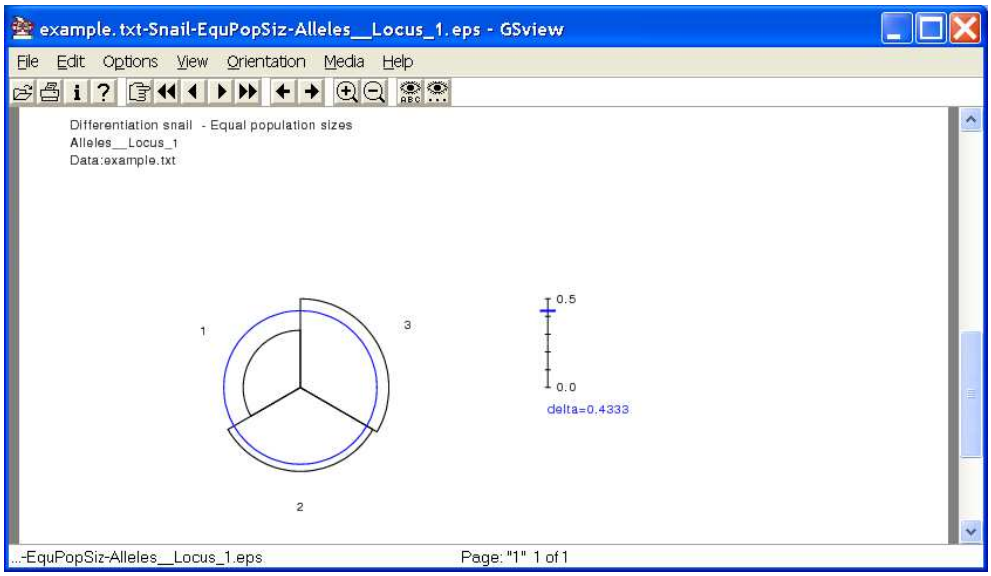


Under "Select one or both methods of calculating relative subpopulation sizes Cj", the first option "Proportional to sample size" is useful if the differing sizes of the demes (e.g. sample, total base population) are to be considered, and "All equal to (1/No. of subpopulations)" if deme size is of no interest (e.g. all underlying populations are considered to be of equal size).

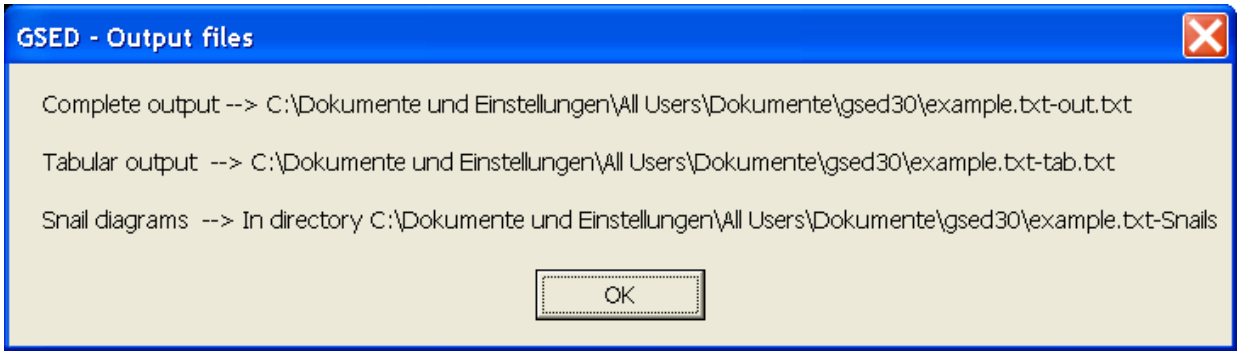
Under "Draw subpopulation differentiation snails in one or both of the formats", checking "WMF (Windows Metafile)" causes each snail to be stored in a vector graphic file in the Metafile format and altered using office programs.



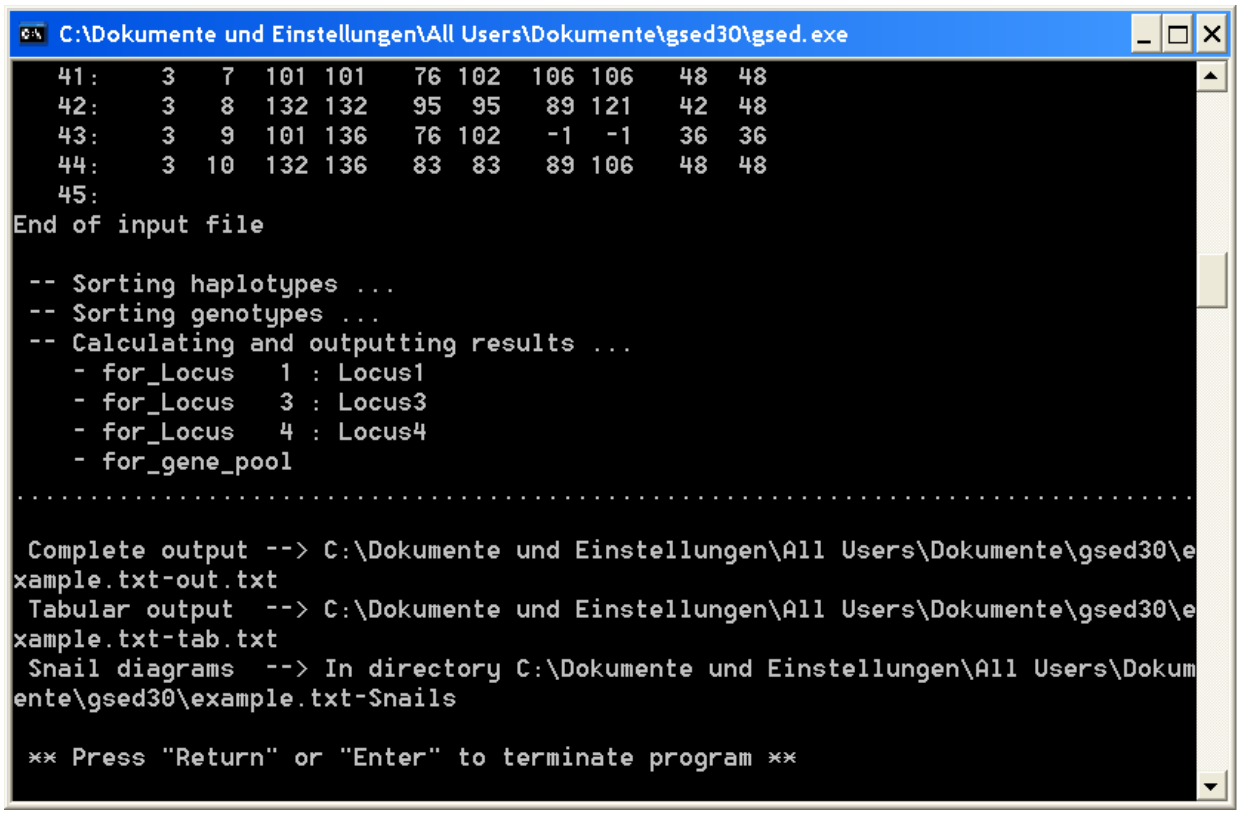
Checking "EPS (Enhanced Postscript)" stores each snail as an eps-file, which can be viewed by programs such as 'ghostview' or 'GSview'.



Upon completion of all calculations, a message window appears that shows the names of the output files.

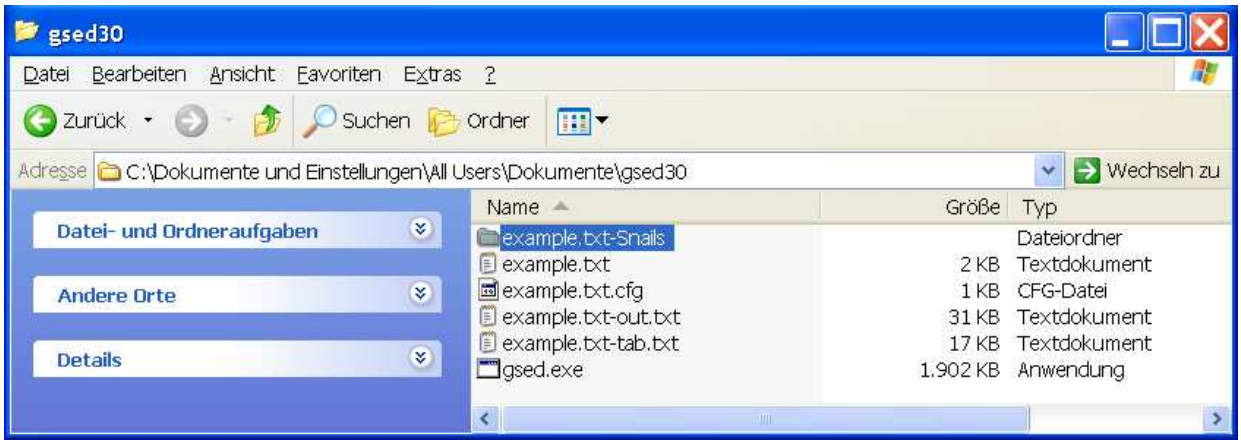


At the same time, the console window contains the following text:



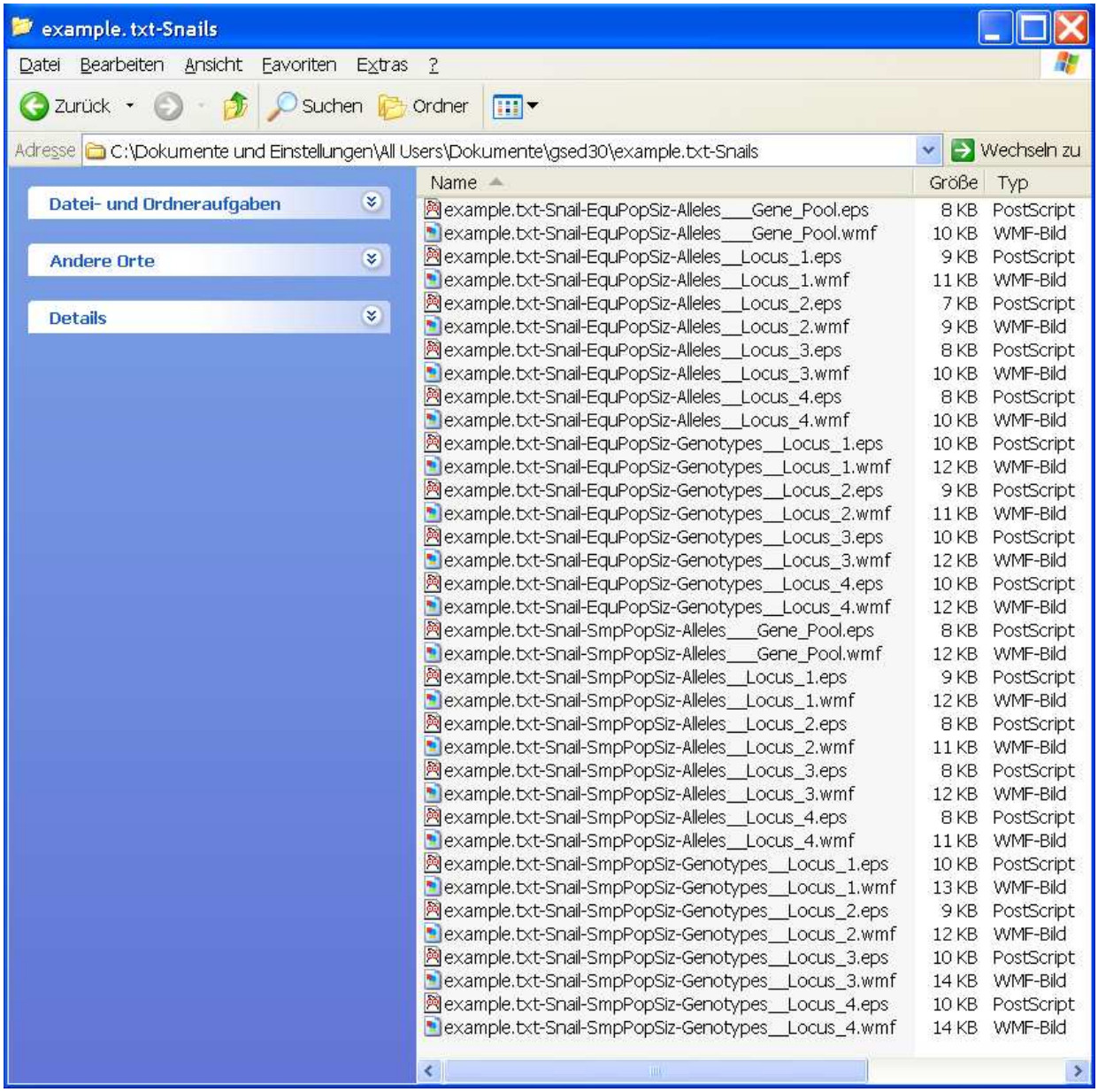
To terminate execution, press “Return” or “Enter” in the console window.

The output files `example.txt-out.txt` and `example.txt-tab.txt` and the directory `example.txt-Snails/` containing the snail files are now listed in the file manager:



The output files contain the following information:

- The file `example.txt-out.txt` contains the complete output, including the results of all statistical tests but excepting the graphical snail files.
- The file `example.txt-tab.txt` contains all of the frequency distributions and all variation parameters in a compact tabular form. The separation of columns by a single tab-character makes this file easily importable into a spreadsheet program.
- The directory `example.txt-Snails/` contains the snail files:



3.2 Keyboard-driven execution

In keyboard-driven execution of GSED, the user answers a sequence of questions through the keyboard. An example is given in App. D.

After specifying a preconstructed input file (see Sec. 2), the user may choose any or all of the frequency distributions listed in Tab. 1 and request calculation of any of the measures and tests listed in Tab. 2. Additional questions concern the format of the output. After all questions have been answered, GSED performs the desired calculations. Results can either be written into an output file or typed on the screen.

Start keyboard-driven execution of GSED by opening a console window and typing

`>gsed -nomenu <Return>`

In the first run of GSED with a new input file, options are chosen by answering a sequence of questions described in more detail in the following (also see App. D). GSED saves the choices in a configuration file that is called up in subsequent runs using this input file (see 3.3).

For questions to be answered by "`Y`" or "`N`", uncapitalized letters "`y`" and "`n`" are also accepted.

```
Enter name of input file (max. 256 characters) ? :
```

Type the name of the input file (see 2), including path specification if necessary. A maximum of 60 characters is allowed.

```
Enter prefix for names of output files [default=example.txt] :
```

Enter a short prefix for the names of all output files. By pressing `<Return>`, the name of the input file is used.

Select the locus configuration:

```
Locus configuration ?     "0" : all single loci    "2" : multilocus - some loci
                          "1" : some single loci   "3" : multilocus - all loci
Option ? :
```

The four options are explained as follows:

```
Option ? : 0
```

Calculations will be carried out for every single locus.

```
Option ? : 1
```

Calculations will be carried out for some of the single loci. As an example, the loci 2 and 4 are specified in reply to the following question:

```
Number of different single loci ? : 2
Which gene loci (separated by commas and using as many lines as necessary) ? :
2,4
Option ? : 2
```

Calculations will be carried out for multilocus genotypes defined by the genotypes at different sets of gene loci, the so-called **multilocus combinations**. As an example, one multilocus combination comprising the gene loci 1 and 2 and a second comprising only the single locus 1 are specified in reply to the following questions (as in the second case, a "multi"locus combination can refer to the genotypes at a single locus):

```
Number of different multilocus combinations ? : 2
Combination 1 :  How many gene loci ? 2
   Which loci (separated by commas and using as many lines as necessary) ? :
1,2
Combination 2 :  How many gene loci ? 1
   Which loci (separated by commas and using as many lines as necessary) ? :
3
```

It is important to note here that measures characterizing the gene pool (see 8) are calculated if and only if the "`Locus configuration ?`" comprises only single loci, *i.e.*, if either option "`0`" or "`1`" is chosen. Since the gene pool measures are formulated as means of the respective single-locus measures at all loci contributing to the gene pool, the single-locus measures must already be available.

```
Option ? : 3
```

Calculations will be performed for multilocus genotypes defined by the genotypes at all of the single loci (in "example.txt" for all four loci).

```
Choice of frequency distributions (Answer "Y" (yes) or "N" (no))
```

Choices can be made among the four types of frequency distribution offered by the subsequent questions and described below (see 5).

```
Should gametic sex specification, if given, be retained ? : Y
```

If no gametic sex is specified at any locus, then the answer to this question is meaningless.

If gametic sex is specified at some or all loci, then an answer of "`N`" will cause this specification to be ignored at all of them. For example, in such a case both of the genotypes A_1A_2 and A_2A_1, where the first allele is that contributed by the maternal parent, would be counted as the genotype A_1A_2.

```
Choice of calculations (Answer "Y" (yes) or "N" (no))
Frequency distributions ? : Y
```

An answer of "Y" causes the calculated frequency distributions to be included in the output. If the answer is "N", they will be omitted.

```
Measures of variation within demes . . .
Measures of variation between demes . . .
Analysis of genotypic structure . . .
```

If selected, the measures and tests offered in the subsequent questions (and described in 6, 7, 8) are calculated for each of the chosen frequency distributions. Note that if `Subpopulation differentiation Dj, delta` is chosen, both methods of calculating relative subpopulation sizes c_j and both of the snail formats WMF and EPS are automatically selected (see 3.1).

The above choices of frequency distributions, measures, and tests are saved in a configuration file. The name of the configuration file consists of the name of the input file followed by ".cfg", for example, "`example.txt.cfg`". All subsequent runs using the same input file will first print the stored configuration table and then ask whether it should be adopted. If the answer is "N", new choices can be made.

An example for the case in which a configuration file already exists is given in App. E.

The sequence of questions continues as follows:

```
Demes for output ?      "0" : all demes        "1" : some demes
Option ? :
```

The two options are explained as follows:

```
Option ? : 0
```

The output contains the results for all of the demes in the input file. Measures of variation between demes (see 6.2) are calculated using ALL of the demes.

```
Option ? : 1
```

The output contains the results for only those demes given in reply to the following question. This option allows measures of variation between demes (see 6.2) to be calculated for differing sets of demes. As an example, demes 1 and 3 are chosen in reply to the following questions:

```
How many demes ? : 2
   Which demes (separated by commas and using as many lines as necessary) ? :
1,3

Output unit ?             "S" : screen             "F" : file
Option ? :
```

Output can be directed to one of two units as follows:

```
Option ? : S
```

All results are typed on the screen. They are not saved elsewhere and thus are lost as soon as they disappear off the screen.

```
Option ? : F
```

Results are output as ASCII text to the designated file. A maximum of 60 characters are allowed for the file name and any necessary specification of path. Since the output is in ASCII code, it is possible to alter its format later using any text editor. The finished file can then be printed on any printer.

```
Output file already exists: "O" : overwrite old output, "P" : enter new prefix :
Option ? :
```

To overwrite the previous contents of the file, which are thus lost, enter the following (note that this option is indicated by the letter "O" and not the numeric character "0" (zero)):

```
Option ? : O
```

To choose a different prefix for the output files, enter the following:

```
Option ? : P
```

If the option is "P", then the following line appears

```
Enter new prefix for output files :
```

in which case one enters a new prefix, such as "xxx". This is followed by

```
Width of output (min. of 75 characters/line) as number of demes per line ? :
     No. demes/line = 1/10 * (No.characters/line - 15)
For example:
     "0"  for ALL   3 demes ( 75 char/line)
     "6"  for       6 demes ( 75 chars/line as for DIN A4 paper upright)
     "10" for      10 demes (115 chars/line as for DIN A4 paper crosswise)
     "11" for      11 demes (125 chars/line as in condensed mode)
Option ? : 0
```

The width of the output medium (e.g., paper) can vary and with it the number of demes that fit onto one line. If the available number of characters per line is known, the formula on the second line above yields the maximal number of demes (rounding down to the nearest integer, if necessary). If not all of the demes fit onto one line, the tables of results are cut off after the specified number of demes and continued on the next lines of output.

The minimum number of characters per line is set to 75. One reason is that this is the length of the commentaries in the output. Another is that the maximal number of columns of the contingency tables in the tests of genotypic structure that can be printed onto one line is also set to the chosen number of demes per line.

A reply of "0" (zero) causes the results of all demes to be printed onto one line.

```
Additional calculations using the same input file and locus configuration ? :
Option ? :
```

When this line appears, the chosen calculations have been completed and either typed on the screen or stored in the output file. Its purpose becomes apparent in the description of the options:

```
Option ? : Y
```

Type "Y" if additional frequency distributions, measures or tests for the same or a different set of demes are desired for the same input file and locus configuration. Since the frequency data is already stored, the input file is not reread and results are obtained quickly. This option provides a means of ordering the output differently from that reflected by the interactive sequence. It also allows calculation of measures of variation between demes for different sets of demes.

```
Option ? : N
```

An answer of "N" terminates the program.

3.3 Configuration file

In subsequent runs of GSED for an input file, a configuration file may exist. This will be the case, if the question "`Should these choices be stored in a file for later use ?`" was answered with "Y" in an earlier run for the same input file. The configuration file contains the previous answers to the questions listed under the headings of "`Choice of frequency distributions`" and "`Choice of calculations`" (see App. D). The name of the configuration file consists of the name of the input file followed by the extension "`.cfg`", for example, "`example.txt.cfg`".

If a configuration file exists, then a configuration table such as that presented in App. E is typed on the screen after the "`Locus configuration`" has been specified.

If the answer to the subsequent question "`Do you want to adopt this configuration ?`" is answered by "`Y`", then "`Choice of frequency distributions`" and "`Choice of calculations`" are skipped. The question "`Should gametic sex specification, if given, be retained ?`" is still posed, since in the case of gametic sex specification, one may want to perform the same calculations with and without regard of gametic sex (see Sec. 5).

If the answer to "`Do you want to adopt this configuration ?`" is "`N`", then "`Choice of frequency distributions`" and "`Choice of calculations`" must be made anew, as in App. E.

3.4 Sorting of haplotypes and genotypes

An answer of "`Y`" to any of the following questions causes the lists of encountered haplotypes and genotypes to be printed in lexicographic order:

```
Frequency distributions ? :
Test of homogeneity of the deme distributions ? :
Test of Hardy-Weinberg structure and heterozygosity ? :
Test of product structure (only if gametic sex is specified) ? :
```

Since sorting of multilocus types can take an extreme amount of computing time, it is advisable not to choose these calculations for multilocus combinations (only the first two questions apply) unless they themselves are of interest. A test of homogeneity for a large number of multilocus types may well exceed the capacity of the program anyway (see A.2). Often, heterozygosity is the only calculation desired for multilocus genotypes; it is performed quickly if it alone is selected.

4 Output

GSED produces four kinds of output, the names of which begin with the chosen prefix (see Sec. 3.1), denoted *prefix*.

The output is named

- The file *prefix*`-out.txt`: Complete output of all selected calculations, including frequency distributions, variation parameters, and statistical tests.
- The file *prefix*`-tab.txt`: Tabular output of all frequency distributions and all calculated variation parameters. This file can be imported into any spreadsheet program by indicating the separation of columns by one tab-character.
- The directory *prefix*`-Snails` containing the snail graphs, if subpopulation differentiation was calculated.
- The file *name of input file*`.cfg`: Stores the selected calculations, to be shown in the checkboxes or configuration of the next run (see Sec. 3.3).

4.1 The output file *prefix*`-out.txt`

This file is organized by locus combination, *i.e.*, by single locus or by multilocus combination (see Secs. 6, 7). See the examples in App. F-I. If calculations are requested only for single loci, results for the gene pool and hypothetical gametic output defined by these loci are included at the end of the file (see Sec. 8).

Each locus combination (single or multilocus) is in turn divided into the output for each of the chosen frequency distributions (measures of variation within demes, between demes) followed by the analysis of genotypic structure (heterozygosity, tests of single locus structure). The demes are listed in columns, as opposed to the file `example.txt-tab.txt`.

Results for measures of variation and heterozygosity appear in tables, each column containing the results for one of the chosen demes. If the chosen width of output (see 3.2: "`Width_of_output ...`") is not sufficient to allow inclusion of all demes onto one line, each table is truncated vertically and continued on the next lines. If the current locus combination consists only of a single locus, the output for this combination closes with the results of the chosen tests of single locus structure for each deme.

The legend printed at the beginning of the output explains notational conventions:

```
O_or_E            Observed_or_Expected_absolute_frequency_in_a_test
{_}               Denotes_multilocus_haplotype_or_genotype
NA                Denotes_undefinable_parameter_value
Gam.sex.spec.?    Abbreviation_of_"Gametic_sex_specification?"
                = "yes",_if_maternal/paternal_alleles_distinguishable
                = "no"_,_otherwise.
alpha             All_alleles/haplotypes/genotypes_of_relative_frequency
                  not_less_than_"alpha"_in_deme_appear_in_sample
                  (with replacement) with_probability>=0.95
alpha-HWP         As_above,_if_genotypes_in_deme_are_in
                  Hardy-Weinberg-Proportions_(HWP)
```

The output for each frequency distribution begins with a heading which provides the following information about each deme:

"Deme_No."	Number of the deme in accordance with the list of demes in the input file.
"Gam.sex.spec.?"	Abbreviation of "Gametic sex specification?" = "yes", if the sex of the parent contributing each allele is known in the entire deme; = "no", otherwise.
"Deme_size"	Total number of individuals whose genotypes are included in the input file, regardless of whether they contain unknown alleles or not.
"No._identified"	Number of individuals whose genetic types with respect to the current frequency distribution are completely identified (no unknown alleles, see 5). Relative frequencies refer to this number.
"alpha"	see below
"alpha-HWP"	see below
"No._unknown"	Number of individuals whose genetic types with respect to the current frequency distribution are unknown and thus are not counted (see 5).

"Deme_size" equals the sum of "No._identified" and "No._unknown".

"alpha": In loose terms, "alpha" tells how frequent a type (allele, haplotype, genotype) must be in the base population in order for it to have a probability of 0.95 or greater of being represented in a deme of the given size ("No._identified"). More precisely, the probability of having sampled (and identified) all types occurring with relative frequency greater than or equal to "alpha" is 0.95 or greater. Obviously, the larger the deme size, the smaller "alpha" becomes. (see Gregorius (1980) for derivation of "alpha".)

Alleles and multilocus haplotypes occur in pairs in the form of genotypes. If the only way to deme haplotypes is by sampling genotypes, it must be remembered that the manner

of association between the different haplotypes making up the genotypes (homozygosity, heterozygosity) has a great influence on the probability of finding the rarer haplotypes. "`alpha`" describes the worst-case situation for finding the rarer haplotypes, namely pure homozygosity, in that only one allele or haplotype can be sampled per individual (see Gregorius (1980) for proof.)

"`alpha-HWP`": In the case of alleles and haplotypes with arbitrary frequencies, this relative frequency characterizes an analogous "`alpha`" for the best-case situation for sampling haplotypes when only genotypes can be sampled. Gregorius (1980) gives proof that this situation occurs when the genotypes arose by random fertilization between alleles/haplotypes, which are then independently associated in the genotypes. The resulting Hardy-Weinberg-Proportions (`HWP`) thus represent the optimal relationships between homozygosity and heterozygosity for sampling different alleles or haplotypes in genotypes. In Gregorius (1980) it is shown that "`alpha-HWP`" is equal to the value of "`alpha`" for a deme twice the size of the given deme. Thus sampling haplotypes in a Hardy-Weinberg population of genotypes is equivalent to drawing a deme of haplotypes singly (as opposed to pairwise) that is twice the size of the given deme of genotypes.

The designation of the different alleles, haplotypes, and genotypes in the output is demonstrated in Tab. 4.

Table 4: *Examples demonstrating designation of alleles, haplotypes and genotypes in the output.*

Genetic type	Designation	Consists of
Allele	`1`	allele 1
Single-locus genotype	`1 x 3`	alleles `1` and `3`
Ordered single-locus genotype	`3 x 1`	maternal allele `3` × paternal allele `1`
Haplotype	`{ 2 4 1 }`	allele `2` at first locus + allele `4` at second locus + allele `1` at third locus
Multilocus genotype	`{ 1 4  2 3  1 3 }`	single-locus genotypes `1 x 4` at first locus + `2 x 3` at second locus + `1 x 3` at third locus
Ordered multilocus genotype	`{ 1 2  4 0  3 2 }`	maternal haplotype `{ 1 4 3 }` × paternal haplotype `{ 2 0 2 }`

The output for the various measures and tests is described in Secs. 6, 7, 8.

The output of the statistical tests is similar to the example in App. F. The upper table contains the observed frequencies of the genotypes A_1A_1, A_1A_3, and A_3A_3 and, beneath each in square brackets, the frequencies expected under the null hypothesis of Hardy-Weinberg structure. The observed allele frequencies are given to the right of the table.

The lower table in App. F, entitled "`Test_statistics`", contains the results of the likelihood ratio test ("`G`"), Pearson's χ^2 test ("`X`"), and, in tables such as this with one degree of freedom ("`DF=1`"), the χ^2 test with continuity correction of 0.5 ("`X**2(c=.5)`"). The symbol (here "`*`") directly to the right of each statistic indicates its level of significance, which can be inferred from the two rightmost columns of the table: The abbreviation "`C.V._of_CHI**2`" stands for "critical value of the χ^2 distribution" for the given degrees of freedom ("`DF`") and "`Level_of_significance`". The symbol "`n.s.`" found to the right of a statistic in other tables means "not significant".

Self-explanatory messages are printed on the screen if difficulties of the following types are encountered: Files cannot be opened, read or closed; an erroneous answer is given during the interactive sequence; limitations on data are exceeded (see A.2). Messages are printed in the output in the following cases: A requested frequency distribution, measure or test cannot be calculated; differences in definition of genetic types between demes prohibit comparison of the demes; special situations arise during a test. Some messages are followed by "`cause=`" and a number, the latter referring to a compiler-specific list of I/O status values.

4.2 The output file *prefix*-tab.txt

The file *prefix*-tab.txt contains all frequency distributions and all calculated variation parameters deme by deme. Its tabular form allows it to be imported into any spreadsheet program by indicating that columns are separated by a single tab-character. See the example in App. G.

The following list relates the abbreviations used in the tabular output file *prefix*-tab.txt with the headings of the output file *prefix*-out.txt:

prefix-out.txt	*prefix*-tab.txt
Deme_No.	Deme
Type_1	Allele or Genotype
DemeSize	No._identified
Alpha	alpha
Alpha-HWP	alpha-HWP
v2-Diver	DIVERSITY_v_2
deltaT	TOTAL_POPULATION_DIFFERENTIATION_delta_T
	EVENNESS_e_FOR_FINITE_POPULATION_SIZE
EvnFinAbs	absolute_e
EvnFinRel	relative_e
EvnFinNum	for_No.types
	EVENNESS_e_FOR_INFINITE_POPULATION_SIZE
EvnInfAbs	absolute_e
EvnInfRel	relative_e
EvnInfNum	for_No.types
	SUBPOPULATION_DIFFERENTIATION_D_j,_delta
	- RELATIVE_SUBPOPULATION_SIZE_PROPORTIONAL_TO_DEME_SIZE
CjDemSiz	Cj=
DjDemSiz	Dj=
deltaDmS	delta=
	- RELATIVE_SUBPOPULATION_SIZES_ALL_EQUAL_TO_(1/NO._SUBPOPS)
CjEquSiz	Cj=
DjEquSiz	Dj=
deltaEqS	delta=

4.3 The output directory *prefix*-Snails

If the calculation of subpopulation differentiation was selected, this directory contains snail graphics of file format "WMF (Windows Metafile)" and/or "EPS (Enhanced Postscript)". See examples in Sec. 3.1.

The names of the snail files are composed by concatenating one element from each of the following sets:

{*prefix*-Snail-} {DemPopSiz, EquPopSiz} {Alleles, Genotypes} {__Locus_x, ___Gene_Pool, __Combination_1__Loci___x__y__z} {.wmf, .eps}

Examples of snail file names are

```
example.txt-Snail-DemPopSiz-Alleles___Gene_Pool.eps
example.txt-Snail-DemPopSiz-Alleles___Gene_Pool.wmf
example.txt-Snail-DemPopSiz-Alleles__Locus_1.eps
example.txt-Snail-DemPopSiz-Alleles__Locus_1.wmf
example.txt-Snail-DemPopSiz-Genotypes__Combination_1__Loci___1__2__3__4.eps
example.txt-Snail-DemPopSiz-Genotypes__Combination_1__Loci___1__2__3__4.wmf
example.txt-Snail-DemPopSiz-Genotypes__Locus_1.eps
example.txt-Snail-DemPopSiz-Genotypes__Locus_1.wmf
example.txt-Snail-EquPopSiz-Alleles___Gene_Pool.eps
example.txt-Snail-EquPopSiz-Alleles___Gene_Pool.wmf
example.txt-Snail-EquPopSiz-Alleles__Locus_1.eps
example.txt-Snail-EquPopSiz-Alleles__Locus_1.wmf
example.txt-Snail-EquPopSiz-Genotypes__Combination_1__Loci___1__2__3__4.eps
example.txt-Snail-EquPopSiz-Genotypes__Combination_1__Loci___1__2__3__4.wmf
example.txt-Snail-EquPopSiz-Genotypes__Locus_1.eps
example.txt-Snail-EquPopSiz-Genotypes__Locus_1.wmf
```

4.4 The configuration file *name of input file*.cfg

This file contains the selected calculations for use in the next run (see Sec. 3.3). It is not necessary to understand this automatically created text file (see Sec. 3.3), which has a form such as

```
 3
yyyyyyy
yy
yy
nnyy
0
yyy
```

5 Frequency distributions

The input to GSED usually consists of the genotypes observed at one or more gene loci in a deme of diploid individuals. (It is also possible to input haplotypes observed in a deme of gametophytes of one sex by listing the second allele at each locus as unknown, *i.e.*, "`-1`", and gametic sex as specified (see 2.2).) From genotype data, it is possible to construct the following frequency distributions:

For a single locus:

- *Allele frequencies:* At a diploid locus, each sampled individual contributes two alleles to the overall deme, so that heterozygotes reveal more allelic types than homozygotes. The association between alleles in genotypes (genotypic structure) therefore determines the degree to which a deme detects the allelic types in a population (see 4: "`alpha`", "`alpha - HWP`").
- *Genotype frequencies:* The genotype of each sampled individual is counted (without regard to gametic sex specification).

Over a set of loci:

- *Multilocus genotype frequencies:* The multilocus genotype of each sampled individual is counted (without regard to gametic sex specification).

Gametic sex specification: In some organisms it is possible to determine which allele at a nuclear gene locus was contributed by the maternal parent. For example, the seed of most coniferous species contains not only the diploid embryo but also nutritive tissue genetically identical to the maternal gametophyte — the primary endosperm or megagametophyte. If the endosperm of a seed is subjected to isoenzyme electrophoresis, the maternal phenotype is revealed. Inheritance analysis of the phenotypes of the endosperm produced by single trees allows inference of the haploid genotype (haplotype) of each endosperm and thus the diploid genotype of each tree ((Bergmann, 1971)). Inference of the genotype of a diploid embryo and "subtraction" of the haplotype of the corresponding endosperm then reveals the haplotype of the paternal gamete for codominant alleles of enzyme loci ((Müller[-Starck], 1977a), (Müller[-Starck], 1977b,Müller[-Starck](@)).

If the gametic sex of the alleles (*i.e.*, the sex of the parent contributing each allele) at all involved loci is specified, a number of additional frequency distributions can be calculated:

For a single locus:

- *Allele frequencies among maternal contributions:* The set of alleles contributed by the maternal parents of the sampled individuals represents a deme of the alleles in the "population" of successful maternal gametes.

- *Allele frequencies among paternal contributions:* In like manner, the set of alleles contributed by the paternal parents of the sampled individuals represents a deme of the alleles in the "population" of successful paternal gametes.

- *Ordered genotype frequencies (maternal × paternal alleles):* The set of ordered genotypes represents a deme out of the "population" of successful fusions between female and male gametes. Ordered genotypes take into account the gametic sex specification of the alleles at the locus, distinguishing for example between the genotypes "`1 x 3`" and "`3 x 1`" (see Tab. 4).

Over a set of loci:

- *Haplotype frequencies among maternal contributions:* The set of maternal haplotypes represents a deme of the haplotypes in the "population" of successful maternal gametes.

- *Haplotype frequencies among paternal contributions:* The set of paternal haplotypes represents a deme of the haplotypes in the "population" of successful paternal gametes.

- *Haplotype frequencies:* A deme of the haplotypes of successful gametes is constructed by counting both the maternal and the paternal haplotypes of the sampled individuals. Since each sampled individual contributes two haplotypes, the association between haplotypes in genotypes (genotypic structure) determines the degree to which a deme detects the haplotypes present in a population (see "*Allele frequencies*" above and see 4: "`alpha`", "`alpha HWP`").

- *Ordered multilocus genotype frequencies (maternal × paternal haplotypes):* The set of ordered genotypes represents a deme of the genotypes in the "population" of successful fusions between female and male gametes. Ordered multilocus genotype frequencies distinguish between maternal and paternal haplotypes. For example, whereas the ordered genotype { `1 2  1 2` } results from fusion of the maternal haplotype { `1 1` } and the paternal haplotype { `2 2` }, the ordered genotype { `2 1  1 2` } is the product of maternal haplotype { `2 1` } and paternal haplotype { `1 2` } (see Tab. 4).

Obtaining unordered genotypes when gametic sex is specified: Note that if gametic sex is specified and the response to the question "`Should gametic sex specification, if given, be retained ?`" is "`Y`" (see 3.2), then the ordered genotype frequency distribution will be calculated, and all measures will be based on this distribution. In order to obtain the unordered distribution and measures calculated for it, GSED must be restarted using the same input file, but a reply of "`N`" must be given to the above question.

Gene pool: If all of the locus combinations that were chosen for calculation were single loci, then the gene pool made up of the genes at these loci is automatically constructed. This will be the case if option "`0`" or "`1`" was given in answer to "`Locus configuration ?`" of

the interactive sequence (see 3.2). Although the frequency distribution of the gene pool is not explicitly included in the output, all of the chosen measures of variation within and between demes are also calculated for the gene pool (see 8) and listed at the end of the output.

Unknown alleles and genotypes: Sometimes it is not possible to determine the genotype or, if gametic sex is specified, one of the parental contributions to an individual at one or more of the investigated loci. In this case, it is up to the user to make sure that the unknown types represent random demes of the respective types in the population. GSED treats unknown alleles (denoted "`-1`" in input) and haplotypes and genotypes containing them as follows for each frequency distribution, :

- *Maternal allele/haplotype frequencies:* Unknown maternal alleles are assumed to be a random deme of all maternal alleles and are thus left out of the calculation. In the same manner, incomplete maternal (multilocus) haplotypes containing an unknown allele at one or more loci are also treated as a random deme of haplotypes and are ignored.
- *Paternal allele/haplotype frequencies:* Unknown paternal alleles and haplotypes are treated in the same way as maternal ones.
- *Allele/haplotype frequencies:* Only those alleles are taken into account that are part of a completely known genotype. Thus if one allele is known and the other is unknown (e.g., because the primary endosperm of a conifer seed was analyzed but the embryo lost), the known allele will not be counted in the allele frequency distribution.
- *Genotype frequencies:* Unknown genotypes are assumed to be a random deme and are not counted.

6 Measures of variation

The following measures of variation can be calculated for any of the types of frequency distributions listed in 5.

6.1 Measures of variation within demes

6.1.1 Diversity v_2

Let a collection be characterized by a frequency vector $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ of its genetic types, where $n \in \mathbb{N}$ and for $k = 1, \ldots, n$ $p_k \geq 0$ and $\sum_{k=1}^{n} p_k = 1$. The **diversity** $v_2(\mathbf{p})$ of the collection is defined as

$$v(\mathbf{p}) = \left(\sum_{k=1}^{n} p_k^2 \right)^{-1}$$

$v_2(\mathbf{p})$ measures the "differentiation effective number" of types; it is less than or equal to the actual number of types and equals this number only for a uniform distribution.

References: Gregorius (1978, 1987)

6.1.2 Total population differentiation δ_T

Let a collection of size N be characterized by a frequency vector $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ of its genetic types, where $n \in \mathbb{N}$ and for $k = 1, \ldots, n$, $p_k \geq 0$ and $\sum_{k=1}^{n} p_k = 1$. The **total population differentiation** δ_T of the collection is defined as

$$\delta_T = \frac{N}{N-1} \cdot \left(1 - \sum_{k=1}^{n} p_k^2 \right)$$

or, letting $N_k := N \cdot p_k$ be the absolute frequency of the kth type,

$$\delta_T = \sum_{k=1}^{n} \frac{N_k}{N} \cdot \frac{N - N_k}{N-1}$$

It holds that $0 \leq \delta_T \leq 1$, with $\delta_T = 0$ for monomorphy and $\delta_T = 1$ if no two deme members are of the same genetic type.

References: Gregorius (1987, 1988)

6.1.3 Evenness e

"Given a distribution of types of individuals in a collection, the **evenness** of the distribution is considered to measure the degree to which these types are equally represented" (Gregorius, 1990).

The **evenness** e is defined to equal one minus the minimal distance of the frequency distribution to all "plateaus", each consisting of equally frequent types, in effectively infinite collections. In small collections, the plateaus are defined by the respective distributions closest to uniformity. If $d_{\min}$ equals this minimal distance, the **absolute evenness** is given by

$$e = 1 - d_{\min}$$

(for the definition of d see "genetic distance" below). $e = 1$ holds only for uniform distributions. As e approaches a lower bound of 0.5, the unevenness increases. As a transformation of evenness which varies between 0 and 1, the **relative evenness** of the population is defined as

$$e = 1 - 2 \cdot d_{\min}$$

Reference: Gregorius (1990)

6.2 Measures of variation between demes

6.2.1 Genetic distance d_0

Let two collections be characterized by frequency vectors $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and $\mathbf{p}' = (p'_1, p'_2 \ldots, p'_n)$ of their genetic types, where $n \in \mathbb{N}$ and for $k = 1, \ldots, n$, $p_k, p'_k \geq 0$ and $\sum_{k=1}^{n} p_k = 1 = \sum_{k=1}^{n} p'_k$. The **genetic distance** $d_0(\mathbf{p}, \mathbf{p}')$ is defined as

$$d_0(\mathbf{p}, \mathbf{p}') = \frac{1}{2} \cdot \sum_{k=1}^{n} |p_k - p'_k|$$

The genetic distance between two collections is specified as the proportion of genetic elements (alleles, genes at multiple loci, gametes, genotypes) which the two collections do not share. Thus $d_0 = 1$ if and only if the two collections have no types in common.

References: Gregorius (1974a,b, 1978, 1984a)

6.2.2 Subpopulation differentiation D_j and δ

Let a population be divided into demes (subpopulations, collections). The amount of **genetic differentiation** of one subpopulation to the remainder of the population is specified as "the proportion of genetic elements (alleles, genes at multiple loci, gametes, genotypes) by which a deme differs from the remainder of the population in type" (Gregorius, 1984b). This proportion is defined as

$$D_j = d_0(\mathbf{p}_j, \bar{\mathbf{p}}_j)$$

where $\mathbf{p}_j$ and $\bar{\mathbf{p}}_j$ are the frequency distributions of the types in deme j and in the remainder of the population, respectively, and d_0 is the genetic distance defined above. The **subpopulation differentiation** is then defined by

$$\delta = \sum_j c_j \cdot D_j$$

where the weights c_j express the proportion of genetic elements present in the jth deme.

References: Gregorius (1984b, 1988, 1996); Gregorius & Roberds (1986)

6.2.3 Test of homogeneity

Let m collections of individuals each be characterized by a frequency distribution defined by the number of individuals of each of n types in the collection. A **test of homogeneity** of the m frequency distributions tests the hypothesis that these m collections all originated from a single large collection of individuals, conditioned on the marginal distributions given by the m deme sizes as proportions of the sum of deme sizes and the mean relative frequencies of the n types over the demes. Goodness-of-fit tests (see 7.2) are performed for $(m-1)(n-1)$ degrees of freedom.

References: Elandt-Johnson (1971, pp.365*ff*), Weber (1978, pp.96*ff*)

7 Analysis of genotypic structure

The following measures and tests aid in the characterization of genotypic structures. In contrast to other measures quantifying variation within and between demes (see 6, 8), heterozygosity measures genetic variation within individuals. Tests of single locus structure investigate the association of gametes in observed (zygotic) genotypic structures by comparing the observed structures to the corresponding expected structures under certain models of association.

7.1 Heterozygosity

7.1.1 Proportion of heterozygosity of single-locus genotypes

Given the genotypes of all individuals in a collection at a single gene locus, the **proportion of heterozygosity** equals the proportion of heterozygous individuals in the collection.

Reference: Gregorius, Krauhausen, Müller-Starck (1986)

7.1.2 Conditional heterozygosity of single-locus genotypes

The **conditional heterozygosity** at a single gene locus takes into account that the proportion of heterozygosity is conditional on the allele frequencies. It results from division of the actual heterozygosity (= proportion of heterozygosity at a single locus) by the corresponding maximum proportion of heterozygosity $H_{\max}$ obtainable for the underlying allele frequencies, where $H_{\max}$ equals 1 if all allele frequencies are less than or equal to 0.5 and $H_{\max} = 2(1 - p)$ if the most frequent allele has frequency p greater than 0.5.

References: Gregorius (1978); Gregorius, Krauhausen, Müller-Starck (1986)

7.1.3 Degree of heterozygosity of multilocus genotypes

The **degree of heterozygosity** is defined for an individual with respect to a specified number of gene loci, and is identical to the proportion of loci at which this individual is heterozygous. The **average degree of heterozygosity** refers to the distribution of this degree in a collection of individuals. Hence it can be proven that the **average degree of heterozygosity** equals the mean proportion of heterozygotes at the single loci.

Reference: Gregorius (1978)

7.2 Tests of single locus structure

The following goodness-of-fit tests are performed for two models of single locus genotypic structure: Pearson's χ^2 **goodness-of-fit test** with statistic

$$X^2 = \sum_{\text{types}} \frac{(N. - E(N.))^2}{E(N.)}$$

Likelihood ratio test with statistic

$$G = 2 \cdot \sum_{\text{types}} N. \cdot (\ln(N.) - \ln(E(N.)))$$

For one degree of freedom, χ^2 **goodness-of-fit test with continuity correction** $\mathbf{c = \frac{1}{2}}$ with statistic

$$X^2_{c=\frac{1}{2}} = \sum_{\text{types}} \frac{\left(|N. - E(N.)| - \frac{1}{2}\right)^2}{E(N.)}$$

$N.$ and $E(N.)$ represent observed and expected deme frequencies, respectively, of the different types.

These statistics are asymptotically χ^2-distributed, the number of degrees of freedom depending on the model. Thus it must be kept in mind that these tests are accurate only for large deme sizes. (A warning is printed in the output if a type is found to have expected frequency less than 5.)

Exact tests have recently been devised in some cases, but these seemed too time-consuming in terms of computing time to allow their inclusion in the larger framework of GSED. In borderline cases (*i.e.*, statistic near critical value of χ^2) of small deme size, it may be advisable to retest structures using special statistics programs that perform exact hypothesis testing.

References: Louis & Dempster (1987), Weir (1990, pp.71*ff*)

7.2.1 Test of Hardy-Weinberg structure and heterozygosity

To each unordered genotypic structure with relative frequencies P_{ij} of genotypes A_iA_j ($P_{ij} = P_{ji}$, $\sum_{i \le j} Pij = 1$) there corresponds a Hardy-Weinberg structure with genotypic frequencies P^*_{ij} defined by

$$P^*_{ii} = p_i^2 \quad \text{and} \quad P^*_{ij} = 2p_ip_j \quad \text{for } i \neq j \text{ and } i, j = 1, \ldots, k$$

In this definition, p_i is the relative frequency of allele A_i from the original genotypic structure, *i.e.*, $p_i = P_{ii} + \frac{1}{2}\sum_{j,j\neq i} P_{ij}$. Hardy-Weinberg structures result from special mating systems, such as are specified, e.g., in Gregorius (1989) pp. 20*ff*, 68*ff*, and Hattemer, Bergmann & Ziehe (1993) pp. 175*ff*.

The purpose here is to detect deviations of (1) an actual genotypic structure (P_{ij}) from its corresponding Hardy-Weinberg structure (P^*_{ij}) and (2) actual heterozygosity from the corresponding Hardy-Weinberg heterozygosity. Actual heterozygosity is defined by $P_{\text{het}} = 1 - \sum_i P_{ii}$ and its corresponding Hardy-Weinberg heterozygosity by $P^*_{\text{het}} = 1 - \sum_i p_i^2$.

Assume that a deme of N individuals was randomly drawn from a large population, and consider their genotypes at a locus with k alleles. (Gametic sex, if specified, is disregarded, *i.e.*, genotypes A_iA_j and A_jA_i are not distinguished.) For unordered absolute genotype frequencies N_{ij} $(i, j = 1, \ldots, k,\ N_{ij} = N_{ji},\ \sum_{i \le j} N_{ij} = N)$ in the deme, the absolute frequency N_i of allele A_i in the deme of $2N$ alleles equals $N_i = 2N_{ii} + \sum_{j \ne i} N_{ij}$. Conditioning on the allele frequencies in the deme (*i.e.*, assuming that the true frequency p_i of allele A_i in the population equals $p_i = N_i/(2N)$), the genotypic frequencies expected under the null hypothesis of Hardy-Weinberg structure equal

$$E(N_{ii}) = N_i^2/(4N) \quad \text{and} \quad E(N_{ij}) = N_iN_j/(2N) \quad (i, j = 1, \ldots, k)$$

The N_{ij} and $E(N_{ij})$ for $i \le j$ are the observed and expected deme frequencies, respectively, entering the test statistics described above. The number of degrees of freedom equals $k(k-1)/2$.

The observed numbers of homozygotes and heterozygotes in a deme of N individuals from a large population equal $\sum_i N_{ii}$ and $\sum_{i<j} N_{ij}$, respectively. The numbers of homozygotes and heterozygotes expected under the assumption of a Hardy-Weinberg structure equal

$$E\left(\sum_i N_{ii}\right) = \left(\sum_i N_i^2\right)/(4N) \quad \text{and} \quad E\left(\sum_{i<j} N_{ij}\right) = N - E\left(\sum_i N_{ii}\right)$$

respectively. Again, the expected frequencies are conditioned on the allele frequencies in the deme. One degree of freedom remains.

By definition, a genotypic structure shows an excess of homozygotes (heterozygotes) if its proportion of homozygotes (heterozygotes) exceeds the proportion of homozygotes (heterozygotes) in the corresponding Hardy-Weinberg structure. If the genotypic structure is a Hardy-Weinberg structure, then such an excess will not be statistically significant; if the test for Hardy-Weinberg structure is not significant, an excess still may or may not be significant. Tests for homozygote excess frequently form the first step in an analysis of so-called "inbreeding structures". Detailed tests for realization of inbreeding structures require consideration of various cases as specified, e.g., in Robertson & Hill (1984).

References: Elandt-Johnson (1971), Gregorius (1989, pp.20*ff*, 68*ff*), Hattemer, Bergmann & Ziehe (1993, pp.175*ff*), Ledwina & Gnot (1980), Pamilo & Varvio-Aho (1984), Robertson & Hill (1984, and references therein), Weir (1990, pp.71*ff*)

7.2.2 Test of product structure for ordered genotypes

In a large population, random fusion of gametes from the set of maternal and the set of paternal gametes gives rise to a zygotic genotypic structure at a locus with k alleles that

fulfills the properties of a product structure

$$P_{ij} = p_i^{\female} p_j^{\male} \quad (i, j = 1, \ldots, k)$$

where P_{ij} is the relative frequency of the ordered genotype A_iA_j (*i.e.*, A_i is the maternal contribution and A_j the paternal, so that $\sum_{i,j} P_{ij} = 1$), $p_i^{\female}$ is the relative frequency of allele A_i among maternal gametic contributions, and $p_i^{\male}$ is the relative frequency of allele A_i among paternal gametic contributions.

Given a random deme of N individuals from a large population, the test of a product structure is performed as a test of independence of association between maternal and paternal allelic contributions conditioned on marginal distributions given by the frequencies of these contributions in the deme. For absolute frequencies N_{ij} $(i, j = 1, \ldots, k)$ of the ordered genotypes in the deme, the absolute frequency $N_i^{\female}$ of the allele A_i in the deme of N maternal alleles equals $N_i^{\female} = \sum_{j=1}^{k} N_{ij}$ and the frequency $N_i^{\male}$ of the allele A_i in the deme of N paternal alleles equals $N_i^{\male} = \sum_{j=1}^{k} N_{ji}$. Conditioning on the allele frequencies in the deme (*i.e.*, assuming that the true frequency $p_i^{\female}$ and $p_i^{\male}$ of allele A_i among the maternal and paternal gametes produced in the population equals $p_i^{\female} = N_i^{\female}/N$ and $p_i^{\male} = N_i^{\male}/N$, respectively), the genotypic frequencies expected under the null hypothesis of a product structure equal

$$E(N_{ij}) = N_i^{\female} N_i^{\male}/N \quad (i, j = 1, \ldots, k)$$

The number of degrees of freedom equals $(k^{\female} - 1)(k^{\male} - 1)$, where $k^{\female}$ and $k^{\male}$ are the numbers of alleles with non-zero frequency among maternal and paternal contributions, respectively.

Reference: Elandt-Johnson (1971, pp.360*ff*)

8 Analysis of the gene pool

"The gene pool of a population with respect to the number [L] of (non-homologous) gene loci located at a certain section of the genome is thought of as the set of all gene (alleles) at these loci realized in all individuals" (Gregorius, 1978)

The following types of gene pool can be constructed, the first two only if gametic sex is specified at all contributing loci:

- Gene pool of maternal contributions
- Gene pool of paternal contributions
- Gene pool of single-locus genotypes

8.1 Measures of variation within demes

8.1.1 Diversity v_2 of the gene pool

Let a collection be characterized at each of L loci by the frequency vector $\mathbf{p}_l = (p_{1l}, p_{2l}, \ldots, p_{nl})$ for $l = 1, \ldots, L$, where $n \in \mathbb{N}$ and for $i = 1, \ldots, n$, $p_{il} \geq 0$ and $\sum_{i=1}^{n} p_{il} = 1$. Denoting by

$$v_{(l)} = \left(\sum_{i=1}^{n} p_{il}^2 \right)^{-1}$$

the allelic diversity at the l-th locus, the **gene pool (genic) diversity** v_2 of the collection was proved to equal the harmonic mean of the single-locus diversities, *i.e.*,

$$v = \left(\frac{1}{L} \cdot \sum_{l=1}^{L} \frac{1}{v_{(l)}} \right)^{-1} = \frac{L}{\sum_{l=1}^{L} \left(\sum_{i=1}^{n} p_{il}^2 \right)}$$

Reference: Gregorius (1987)

8.1.2 Diversity v_{gam} of the hypothetical gametic output

Let a collection be characterized at locus l ($l = 1, \ldots, L$) by the frequency vector $\mathbf{p}_l = (p_{1l}, p_{2l}, \ldots, p_{nl})$, where $n \in \mathbb{N}$ and for $i = 1, \ldots, n$, $p_{il} \geq 0$ and $\sum_{i=1}^{n} p_{il} = 1$. Denoting by

$$v_{(l)} = \left(\sum_{i=1}^{n} p_{il}^2 \right)^{-1}$$

the allelic diversity at the l-th locus, the **hypothetical gametic diversity** v_{gam} of the collection is defined as

$$v_{gam} = \prod_{l=1}^{L} v_{(l)}$$

The hypothetical gametic output is defined by the set of gametes that results from stochastically independent association between loci, free recombination, and equal gametic production for all members. v_{gam} therefore measures the potential of a population for producing genetically diverse gametes.

Reference: Gregorius (1978)

8.1.3 Total population differentiation δ_T of the gene pool

Let a collection of subpopulations have the total population differentiation $\delta_{T(l)}$ at locus l $(l = 1, \ldots, L)$. Then the **total population differentiation δ_T of the gene pool** was proven to equal the arithmetic mean of the total population differentiation at each locus, that is,

$$\delta_T = \frac{1}{L} \cdot \sum_{l=1}^{L} \delta_{T(l)}$$

Reference: Gregorius (1987)

8.2 Measures of variation between demes

8.2.1 Distance d_0 between gene pools

Let one collection be characterized by the frequency vectors of the different genes (alleles) at L gene loci, that is, by the L frequency vectors $\mathbf{p}_l = (p_{1l}, p_{2l}, \ldots, p_{n_l l})$ $(l = 1, \ldots, L)$, where $n_l \in \mathbb{N}$ is the number of alleles at locus l and $p_{kl} \geq 0$ and $\sum_{k=1}^{n_l} p_{kl} = 1$ holds for all $k = 1, \ldots, n_l$. Let a second collection be characterized by the L frequency vectors $\mathbf{p}'_l = \left(p'_{1l}, p'_{2l}, \ldots, p'_{n_l l}\right)$ $(l = 1, \ldots, L)$ at the same L loci and for the same numbering of alleles at each locus. The **gene pool genetic distance** d_0 between the two collections was proven to be the arithmetic mean of the single-locus distances, *i.e.*,

$$\begin{aligned} d_0 &= \frac{1}{L} \cdot \sum_{l=1}^{L} d_0(\mathbf{p}_l, \mathbf{p}'_l) \\ &= \frac{1}{L} \cdot \sum_{l=1}^{L} \left(\frac{1}{2} \cdot \sum_{k=1}^{n_l} |p_{kl} - p'_{kl}| \right) \end{aligned}$$

Reference: Gregorius & Roberds (1986)

8.2.2 Differentiation δ of subdivided gene pools

Let a collection of subpopulations have the subpopulation differentiation $\delta_{(l)}$ at locus l $(l = 1, \ldots, L)$. Then the (unweighted) **subpopulation differentiation δ of the gene**

pool was proven to be the arithmetic mean of the subpopulation differentiation at each locus, that is

$$\delta = \frac{1}{L} \cdot \sum_{l=1}^{L} \delta_{(l)}$$

Reference: Gregorius & Roberds (1986)

9 Acknowledgements and disclaimer

I am still grateful to the many colleagues who helped with the original version of GSED and its manual. Matthias Kühle did some of the programming of the interactive sequence, especially the saving of the configuration. Fritz Bergmann, Bernd Degen, Reiner Finkeldey, Hans-Rolf Gregorius, Hans H. Hattemer, Sven Herzog, Bernhard Hosius, Gerhard Müller-Starck, Aristotelis Papageorgiou, Rommy Starke, Jozef Turok, Martin Ziehe, and too many master's students to list here tested the various versions of GSED on their data and suggested improvements. Martin Ziehe recalculated many of the computed results, thereby discovering several bugs. Hans-Rolf Gregorius provided valuable instruction over the years on the meaning of genetic variation and mating systems in general and on the implemented measures and tests in particular and suggested improvements for the output. Hans-Rolf Gregorius, Hans H. Hattemer, Bernhard Hosius, and Martin Ziehe suggested improvements on this manual.

Generous financial support of earlier versions was obtained by Florian Scholz† and Bernd Degen of the Bundesforschungsanstalt für Forst- und Holzwirtschaft (now the Institut für Forstgenetik in the Johann Heinrich von Thünen-Institut) in Großhansdorf and Alwin Janßen of the Hessische Forstliche Versuchsanstalt (now the Nordwestdeutsche Forstliche Versuchsanstalt) in Hann. Münden.

I am also grateful to Helmut Michels, author of the scientific data plot software DISLIN http://www.dislin.de, for advice on programming the menu-driven execution and on plotting the subpopulation differentiation snails in the new version of GSED.

As before, I have tried my best to find all programming errors. Nevertheless, the user is advised to check the correctness of the results, as I can assume no liability for any errors. I would be very grateful for news of any bugs that remain in the program or errors in this manual.

10 Copyright information on the GSED software

Author's address:
Abt. Forstgenetik und Forstpflanzenzüchtung
Universität Göttingen
Büsgenweg 2
37077 Göttingen, Germany

Email: egillet@gwdg.de

GSED website: http://www.uni-goettingen.de/de/67064.html

References

Bergmann, F. 1971. Genetische Untersuchungen bei *Picea abies* mit Hilfe der Isoenzym-Identifizierung. II. Möglichkeiten für genetische Zertifizierung von Forstsaatgut. *Allgemeine Forst- und Jagdzeitung* 142, 278-280.

Elandt-Johnson, R.C. 1971. *Probability Models and Statistical Methods in Genetics.* John Wiley & Sons, Inc., New York, London, Sydney, Toronto.

Emigh, T.H. 1980. A comparison of tests for Hardy-Weinberg equilibrium. *Biometrics* 36, 627-642. http://www.jstor.org/stable/2556115

Gillet, E.M. 1996. Qualitative inheritance analysis of isoenzymes in haploid gametophytes: Principles and a computerized method. *Silvae Genetica* 45, 8-16. http://www.bfafh.de/inst2/sg-pdf/45_1_8.pdf

Gregorius, H.-R. 1974a. Genetischer Abstand zwischen Populationen. I. Zur Konzeption der genetischen Abstandsmessung. *Silvae Genetica* 23, 22-27. http://www.bfafh.de/inst2/sg-pdf/23_1-3_22.pdf

Gregorius, H.-R. 1974b. On the concept of genetic distance between populations based on gene frequencies. *Proc. Joint IUFRO Meeting, S02.04.1-3, Stockholm, Session I*, 17-26.

Gregorius, H.-R. 1978. The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. *Math. Biosciences* 41, 253-271.http://dx.doi.org/10.1016/0025-5564(78)90040-8

Gregorius, H.-R. 1980. The probability of losing an allele when diploid genotypes are sampled. *Biometrics* 36, 643-652. http://www.jstor.org/stable/2556116

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

Gregorius, H.-R. 1984b. Measurement of genetic differentiation in plant populations. Pp. 276-285 in: Gregorius, H.-R. (ed.). *Population Genetics in Forestry.* Springer-Verlag, Berlin, Heidelberg, New York, Tokyo.

Gregorius, H.-R., Roberds, J.H. 1986. Measurement of genetical differentiation among subpopulations. *Theor. Appl. Genet.* 71, 826-834. http://dx.doi.org/10.1007/BF00276425

Gregorius, H.-R. 1987. The relationship between the concepts of genetic diversity and differentiation. *Theor. Appl. Genet.* 74, 397-401. http://dx.doi.org/10.1007/BF00274724

Gregorius, H.-R. 1988. The meaning of genetic variation within and between subpopulations. *Theor. Appl. Genet.* 76, 947-951. http://dx.doi.org/10.1007/BF00273686

Gregorius, H.-R. 1989. *Characterization and Analysis of Mating Systems.* Ekopan Verlag, Witzenhausen. http://webdoc.sub.gwdg.de/ebook/y/2001/gregorius/matesys.pdf

Gregorius, H.-R. 1990. A diversity-independent measure of evenness. *American Naturalist* 136, 701-711. http://dx.doi.org/10.1086/285124

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. 2009. Distribution of variation over populations. *Theory in Biosciences* 128, 179-189. http://dx.doi.org/10.1007/s12064-009-0064-1

Gregorius, H.-R., Krauhausen, J. & Müller-Starck, G. 1986. Spatial and temporal genetic differentiation among the seed in a stand of *Fagus sylvatica* L. *Heredity* 57, 255-262. http://dx.doi.org/10.1038/hdy.1986.116

Hattemer, H.H, Bergmann, F. & Ziehe, M. 1993. *Einführung in die Genetik für Studierende der Forstwissenschaft*, 2. Aufl. J.D. Sauerländer's Verlag, Frankfurt am Main.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* 17, 4015-4026. http://dx.doi.org/10.1111/j.1365-294X.2008.03887.x

Kim, Z.-S. 1985. Viability selection at an allozyme locus during development in European beech (*Fagus sylvatica* L.) *Silvae Genetica* 34, 181-186. http://www.bfafh.de/inst2/sg-pdf/34_4-5_181.pdf

Ledwina, T., Gnot, S. 1980. Testing for Hardy-Weinberg equilibrium. *Biometrics* 36, 161-165. http://www.jstor.org/stable/2530507

Louis, E.J., Dempster, E.R. 1987. An exact test for Hardy-Weinberg and multiple alleles. *Biometrics* 43, 805-811. http://www.jstor.org/stable/2531534

Michels, H. 2009. DISLIN Scientific Plotting Software. ©Helmut Michels, Max-Planck-Institut für Sonnensystemforschung, Katlenburg-Lindau. http://www.dislin.de.

Müller[-Starck], G. 1977a. Untersuchungen über die natürliche Selbstbefruchtung in Beständen der Fichte (*Picea abies* (L.) Karst.) und Kiefer (*Pinus sylvestris* L.). *Silvae Genetica* 26, 207-217. http://www.bfafh.de/inst2/sg-pdf/26_5-6_207.pdf

Müller[-Starck], G. 1977b. Cross-fertilization in a conifer stand inferred from enzyme gene markers in seeds. *Silvae Genetica* 26, 223-226. http://www.bfafh.de/inst2/sg-pdf/26_5-6_223.pdf

Pamilo, P.,Varvio-Aho, S. 1984. Testing genotype frequencies and heterozygosities. *Marine Biology* 79, 99-100. http://dx.doi.org/10.1007/BF00404990

Robertson, A., Hill, W.G. 1984. Deviations from Hardy-Weinberg proportions: Sampling variances and use in estimation of inbreeding coefficients. *Genetics* 107, 703-718. http://www.genetics.org/cgi/content/abstract/107/4/703

Weber, E. 1978. *Mathematische Grundlagen der Genetik.* VEB Gustav Fischer Verlag, Jena.

Weir, B.S. 1990. *Genetic Data Analysis.* Sinauer Associates, Inc. Publ., Sunderland, Mass.

Wright, S. 1978. *Evolution and the Genetics of Populations.* Vol. 4. The University of Chicago Press, Chicago.

A Technical specifications

A.1 Compiler information

GSED is written in the programming language FORTRAN as a mixture of FORTRAN-77 and FORTRAN-90 subroutines. Compilations by the GNU FORTRAN compiler `gfortran` on the operating systems WinXP and openSUSE Linux can be downloaded from http://www.uni-goettingen.de/de/95607.html. The compiler version appears in the program header.

During execution, GSED stores intermediate results in direct access files. Depending on the compiler they receive names such as "FORnnn.DAT" or a seemingly arbitrary sequence of letters and numbers. They are stored on the default directory or drive (see A.1) and are deleted automatically upon successful completion of the program. If the program is interrupted in mid-run, the files may remain and can be deleted by hand.

A.2 Limitations on data

Allele designations, or "names", must be non-negative integers. No non-numeric letters are permitted. The allele designation "0" is meant to denote a "null allele", if its presence can be determined.

All integers are of type INTEGER(4) and range between -2147483647 and +2147483647. Real numbers are of single precision type REAL with approximately 7 digit accuracy and range from 10^{-38} to 10^{38}.

The output formats accomodate 5-digit integers (up to 99999) and floating point numbers with up to 5 digits in front of the decimal point. Floating point calculations are printed with 3 decimal places. The one exception is the expected absolute frequencies in tests, which have two decimal places.

GSED currently allows a maximum of `MXALL`=250 different allele designations, or "names", across loci. For example, if alleles "200", "201", and "202" are found at locus 1 over all demes, and if "200", "210", and "220" occur at locus 2 in all demes, then there are five designations, namely "200", "201", "202", "210", and "220".

Statistical tests can be performed for a maximum of 100 degrees of freedom. An encounter of more degrees of freedom does not cause termination of the program. If the expected value in any cell is less than 5, a warning is printed. Smaller expected values can inflate the X^2 and G test statistics, resulting in the erroneous indication of significant deviation from the hypothesis as compared to an exact test.

B The example input file `example.txt`

The following sample input file is used throughout the manual. The genotypes at all 4 loci are unordered, *i.e.*, lacking gametic specification. Note that individual 9 at Locus 3 in Deme 3 is of unknown genotype ("`-1 -1`"). The last line of the file is an empty line.

```
3   4
Locus1
Locus2
Locus3
Locus4
Pop1
Pop2
Pop3
unformatted
1    0     1     0    2    0     3     0    4    0
1    1   121   121   83   95   118   193   42   42
1    2   121   136   83  102   106   193   36   36
1    3   136   136   76   95   118   121   42   48
1    4   101   136   95  102   106   121   42   48
1    5   121   136   83  102   106   118   36   42
1    6   101   101   83  102   106   193   36   36
1    7   101   136   83   83   106   193   42   48
1    8   101   121   95   95    89   106   42   48
1    9   121   121   76   83    89   193   36   36
1   10   101   121   76   95   118   121   42   42

2    0     1     0    2    0     3     0    4    0
2    1   101   101   76   83    89   106   36   48
2    2   101   101   83   83   106   106   36   48
2    3   101   101   76   76   121   193   36   48
2    4   121   121   83   83    89    89   36   42
2    5   121   121   76   76   121   193   36   42
2    6   101   121   83   95   106   193   36   48
2    7   101   121   83  102    89   193   36   48
2    8   101   121   76   95    89    89   36   48
2    9   101   121   76  102    89   193   36   48
2   10   101   121   76  102   121   121   36   48

3    0     1     0    2    0     3     0    4    0
3    1   132   136   76   83    89   106   36   42
3    2   101   101   76   83   106   106   42   42
3    3   132   136   95   95   121   121   36   48
3    4   101   136   83   95   121   121   36   42
3    5   101   136   83  102    89    89   36   42
3    6   132   132   76   76    89    89   42   42
3    7   101   101   76  102   106   106   48   48
3    8   132   132   95   95    89   121   42   48
3    9   101   136   76  102    -1    -1   36   36
3   10   132   136   83   83    89   106   48   48
```

C More examples of input files

Example 1: The file "example.txt" was already introduced (see Sec. 2, 3).

The first deme ("`Population 1`") consists of the 4-locus genotypes of 5 individuals designated "`1`", "`3`", "`4`", "`5`", and "`6`"; gametic sex is specified at all loci.

"`Population 2`" consists of the 4-locus genotypes of the 5 individuals "`1`"–"`5`"; gametic sex is specified at loci 1 and 2 but not loci 3 and 4.

In "`Population 3`", nine different genotypes were found in a deme of 100 individuals, the frequencies of the different genotypes in the deme equalling 32, 19, ...; gametic sex is not specified at any locus. This constellation of gametic sex specification may not be very realistic, but it demonstrates the form of data input and in particular the meaning of the key line (see 2.3.1).

```
   3   4
LAP-A
LAP-B
IDH
PGI
Population 1
Population 2
Population 3
unformatted
   1   0  1 1  2 1  3 1  4 1
   1   1  1 3  1 2  3 3  2 3
   1   3  3 1  1 1  3 3  2 2
   1   4  3 1  1 1  3 3  2 2
   1   5  1 1  1 1  3 3  2 2
   1   6  3 3  2 1  1 3  3 3

   2   0  1 1  2 1  3 0  4 0
   2   1  3 3  1 2  1 1  2 3
   2   2  1 3  2 1  1 3  2 3
   2   3  3 1  1 1  3 3  2 3
   2   4  1 1  1 1  1 3  2 2
   2   5  3 3  1 2  3 3  2 3

   3   0 -1 0  2 0  3 0  4 0
   3  32  3 3  1 2  1 1  2 3
   3  19  1 3  1 2  1 3  2 3
   3   4  1 3  1 1  3 3  2 3
   3   7  1 1  1 1  1 3  2 2
   3  25  3 3  1 2  3 3  2 3
   3   4  1 3  1 1  1 3  2 3
   3   3  3 3  1 2  3 3  2 3
   3   3  1 3  1 1  1 3  3 3
   3   3  1 3  1 2  1 3  2 3
```

Example 2: This example shows microsatellite data, in which the qdesignation of each allele correspond to its number of amplified base pairs. In fact, the allele designation in data can be any positive integer. The allele designation "-1" indicates missing data. Gametic sex is not specified, as indicated by the "0" following each locus number in the deme specification line.

```
2 5
Loc1
Loc2
Loc3
Loc4
Loc5
Pop1
Pop2
unformatted
1     0       1   0      2   0      3   0      4   0      5   0
1     1     213 223    244 244    154 168    155 155    191 193
1     2     213 217    244 244    166 168    155 155    187 193
1     3     213 215    243 248    166 168    155 155    187 187
1     4     211 211    240 242    166 168    151 155    187 201
1     5     213 217    243 248    166 168    155 155    187 193
1     6     211 223    243 243    168 168    151 155    187 193
1     7     211 215    243 245    168 168    155 173    187 203
1     8      -1  -1    248 251    154 168    155 173    197 197
1     9     215 217    243 243    154 168    155 159    191 191
1    10     211 213    255 255    166 168    155 155    187 201

2     0       1   0      2   0      3   0      4   0      5   0
2    65     213 223    244 246    166 168    155 155    187 197
2    66     213 215    243 243    168 168    155 155    187 187
2    67     215 225    242 246    168 168    155 159    201 201
2    68     213 225    242 245    168 168    155 173    189 201
2    69     217 229    246 252    168 168    155 157    187 187
2    70     203 215    242 246    168 168    155 155    187 197
2    71     213 219    243 246    168 168    155 165    191 197
2    72     213 235    240 244    168 168    155 161    191 197
2    73     209 209    244 244    168 168    155 155    187 205
2    74     203 229    242 242    166 168    155 161    187 193
2    75     213 213    242 244    168 168    153 155    187 197
```

Example 3: This example of formatted input (see 2.2.2) demonstrates that it is still possible to input single locus genotypes at more than one locus, even if they refer to the same population but not necessarily the same individuals. The disadvantage of this type of input is that all individual multilocus information is lost. The "-1" in the deme specification lines indicates that the second field of each genotype line contains the number of individuals that were found to have the respective genotype in the deme. Note that it is possible to include several deme specification lines within the same deme. (Data adapted from Kim (1985))

```
   2   2
SAP-A
LAP-A
DEUTSCHLAND: ECKERN
DEUTSCHLAND: KEIMLINGE GEWAECHSHAUS
 1  (2I4,1X,1(2I2,1X))
   1   0 -2 0
   1 151  1 1
   1 111  2 2
   1 107  3 3
   1   6  4 4
   1  51  1 2
   1  83  1 3
   1   5  1 4
   1  68  2 3
   1   7  2 4
   1   3  3 4
9999
   2   0 -1 0
   2  23  1 1
   2  71  2 2
   2   1  3 3
   2  62  1 2
   2   2  1 3
   2   3  1 4
   2   6  2 3
   2   6  2 4
   2   0 -2 0
   2  39  1 1
   2  53  2 2
   2  32  3 3
   2   2  4 4
   2   6  1 2
   2   9  1 3
   2   1  1 4
   2  16  2 3
   2   2  2 4
   2   2  3 4
9999
```

Example 4: Here the number of loci is large, necessitating continuation of the genotypes on additional lines. The usage of the formatted input saves space on the line by allowing the alleles designation to be written with separators. Lack of gametic sex specification is indicated by the "0" following each locus number in the deme specification line.

```
   1  30
Locus  1
Locus  2
...
Locus 30
Beech forest
30  (2i4,20(i2,i1)/8x,10(i2,i1))
   1   0 10 20 30 40 50 60 70 80 90100110120130140150160170180180200
        210220230240250260270280290300
   1   1 11 22 31 11 21 12 44 34 35 43 31 32 33 22 13 13 22 11 11 22
         43 23 22 33 11 22 31 12 11 00
   1   2 22 33 33 22 44 11 22 11 31 11 22 31 12 14 34 42 13 23 21 22
         35 43 21 12 11 23 21 11 22 12
9999
```

Example 5: This example of formatted input shows the input for a deme of "successful" maternal haplotypes, such as could be found by sampling the bulk seed of a stand of a conifer species and subjecting only the primary endosperm of each seed to isoenzyme electrophoresis (see 5). The specification of gametic sex is indicated by the "1" following each locus number in the deme specification line. The unknown paternal contribution at each locus is designated by "-1".

```
   1   4
Locus 1
Locus 2
Locus 3
Locus 4
Scots pine forest
 4  (2i4,4(i3,i2))
   1   0  1 1  2 1  3 1  4 1
   1   1  2-1  3-1  3-1  1-1
   1   2  2-1  2-1  1-1  2-1
   1   3  1-1  3-1  2-1  3-1
   1   4  1-1  2-1  1-1  2-1
9999
```

D Keyboard-driven execution for first run

This example of keyboard input shows the first run of GSED for an input file named "example.txt".

(GSED heading)

```
 Enter name of input file (max. 256 characters) ? : example.txt
 Enter prefix for names of output files [default=example.txt] : <Return>
 ..........................................................................
Locus  1: LAP-A
Locus  2: LAP-B            Deme  1: Population 1
Locus  3: IDH              Deme  2: Population 2
Locus  4: PGI              Deme  3: Population 3
 ..........................................................................
 Locus configuration ?     "0" : all single loci    "2" : multilocus - some loci
                           "1" : some single loci   "3" : multilocus - all loci
 Option ? : 0
 ..........................................................................
 Choice of frequency distributions (Answer "Y" (yes) or "N" (no))
    Allele/haplotype frequencies among maternal contributions ? : N
    Allele/haplotype frequencies among paternal contributions ? : N
    Allele/haplotype frequencies ? : Y
    Genotype frequencies ? : Y
 ..........................................................................
 Ignore gametic sex, if specified in data ? N
 ..........................................................................
 Choice of calculations (Answer "Y" (yes) or "N" (no))

    Frequency distributions ? : Y

    Measures of variation within demes
    ----------------------------------
    Diversity v_2 ? : Y
    Total population differentiation  delta_T? : Y
    Evenness : Y
      -    finite population size ? : Y
      -  infinite population size ? : Y

    Measures of variation between demes
    -----------------------------------
    Genetic distance d_0 ? : Y
    Subpopulation differentiation D_j, delta : Y
      -  subpopulations weighted proportional to deme size ? : Y
      -  subpopulations equally weighted ? : Y
    Test of homogeneity of the deme distributions ? : Y

    Analysis of genotypic structure
    -------------------------------
    Heterozygosity ? : Y
    Tests of single locus structure :
       Test of Hardy-Weinberg structure and heterozygosity ? : Y
       Test of product structure (only if gametic sex is specified) ? : Y

Options saved in file example.txt.cfg
 ..........................................................................
```

```
Deme  1: Population 1
Deme  2: Population 2
Deme  3: Population 3

Demes for output ?      "0" : all demes        "1" : some demes
Option ? : 0
.....................................................................
Output unit ?             "S" : screen             "F" : file
Option ? : F
.....................................................................
Width of output (min. of 75 characters/line) as number of demes per line ? :
     No. demes/line = 1/10 * (No.characters/line - 15)
For example:
     "0"  for ALL   3 demes ( 75 char/line)
     "6"  for       6 demes ( 75 chars/line as for DIN A4 paper upright)
     "10" for      10 demes (115 chars/line as for DIN A4 paper crosswise)
     "11" for      11 demes (125 chars/line as in condensed mode)
Option ? : 0
.....................................................................
Reading input file example.txt
   1: 3        4
   2: Locus1
   3: Locus2
   4: Locus3
   5: Locus4
   6: Pop1
   7: Pop2
   8: Pop3
   9: unformatted
  10:    1   0    1   0    2   0    3   0    4   0
  11:    1   1  121 121   83  95  118 193   42  42
...
  44:    3  10  132 136   83  83   89 106   48  48
  45:
End of input file

-- Sorting haplotypes ...
-- Sorting genotypes ...
-- Calculating and outputting results ...
   - for locus No.    1:  Locus1
   - for locus No.    2:  Locus2
   - for locus No.    3:  Locus3
   - for locus No.    4:  Locus4
.....................................................................
Additional calculations using the same input file and locus configuration ? :
Option ? : N
```

E Keyboard-driven execution for subsequent runs

Start of the interactive sequence for a subsequent run using the input file "example.txt". The choices made during the first run were stored in the configuration file `example.txt.cfg` shown below. This configuration can be adopted by replying with a "Y", in which case the interactive sequence will be skipped. A reply of "N" allows a new choice of frequency distributions and calculations.

```
 Enter name of input file (max. 256 characters) ? :
example.txt
 Enter prefix for names of output files [default=example.txt] :
<Return>

 Option file exists : example.txt.cfg
 ------------------------------------------------------------------------------
 | Frequency distributions                  y | Maternal frequencies         n|
 |.....MEASURES.OF.VARIATION.WITHIN.DEMES.... | Paternal frequencies         n|
 | Diversity                                y | Allele/haplotype frequencies y|
 | Total population differentiation deltaT  y | Genotype frequencies         y|
 | Evenness                                 y ---------------------------------|
 |  - finite population size                y                                  |
 |  - infinite population size              y                                  |
 |....MEASURES.OF.VARIATION.BETWEEN.DEMES....                                  |
 | Genetic distance                         y                                  |
 | Subpopulation differentiation Dj, delta  y                                  |
 |  - weights proportional to sample size   y                                  |
 |  - weights all equal to (1/No. subpops)  y                                  |
 | Test of homogeneity                      y                                  |
 |.................ANALYSIS.OF.GENOTYPIC.STRUCTURE...................         |
 | Heterozygosity                                                 y        |
 | Test of Hardy-Weinberg structure + heterozygosity (single locus)  y        |
 | Test of product structure                        (single locus)  y        |
 ------------------------------------------------------------------------------
 Do you want to use these options ? Answer "Y" (yes) or "N" (no):
y
.....................................................................
Locus  1: Locus1
Locus  2: Locus2          Deme  1: Pop1
Locus  3: Locus3          Deme  2: Pop2
Locus  4: Locus4          Deme  3: Pop3
.....................................................................
 Locus configuration ?   "0" : all single loci    "2" : multilocus - some loci
                         "1" : some single loci   "3" : multilocus - all loci
 Option ? :
0
.....................................................................
 Ignore gametic sex, if specified in data ?
n
Options saved in file example.txt.cfg
.....................................................................
Deme    1 : Pop1
Deme    2 : Pop2
Deme    3 : Pop3

 Demes for output ?      "0" : all demes        "1" : some demes
 Option ? :
0
.....................................................................
 Output unit ?             "S" : screen             "F" : file
```

```
 Option ? :
f

 File already exists:      "A" : append new output  "O" : overwrite old output
 Option ? :
o
.....................................................................
 Width of output (min. of 75 characters/line) as number of demes per line ? :
      No. demes/line = 1/10 * (No. characters/line - 15)
 For example:
      "0"  for ALL   3 demes ( 75 chars/line)
      "6"  for       6 demes ( 75 chars/line as for DIN A4 paper upright)
      "10" for      10 demes (115 chars/line as for DIN A4 paper crosswise)
      "11" for      11 demes (125 chars/line as in condensed mode)
 Option ? :
0
.....................................................................
Reading input file example.txt
    1: 3        4
    2: Locus1
    3: Locus2
    4: Locus3
    5: Locus4
    6: Pop1
    7: Pop2
    8: Pop3
    9: unformatted
   10:    1   0    1   0    2   0    3   0    4   0
   11:    1   1  121 121   83  95  118 193   42  42
...
   44:    3  10  132 136   83  83   89 106   48  48
   45:
End of input file

 -- Sorting haplotypes ...
 -- Sorting genotypes ...
 -- Calculating and outputting results ...
    - for_Locus   1 : Locus1
    - for_Locus   2 : Locus2
    - for_Locus   3 : Locus3
    - for_Locus   4 : Locus4
    - for_gene_pool
.....................................................................
 Additional calculations using same input file and locus configuration ? :
 Option ? :
n
.....................................................................
 Complete output --> example.txt-out.txt
 Tabular output  --> example.txt-tab.txt
 Snail diagrams  --> In directory example.txt-Snails

 ** Press "Return" or "Enter" to terminate program **
<Return>
```

F Output file `example.txt-out.txt`

Beginning of the output file for single-locus genotypes in the sample input file `example.txt-out.txt`

```
*************************************************************************
Locus   1 - Locus1
*************************************************************************

=========================================================================
Allele_frequencies_among_individuals_of_identified_genotype
=========================================================================
Deme                1         2         3
Gam.sex.spec.?     no        no        no
Deme_size          10        10        10
No._identified     10        10        10
alpha           0.334     0.334     0.334
alpha-HWP       0.201     0.201     0.201
No._unknown         0         0         0

-----------------------------------
Absolute_frequency_distribution
-----------------------------------
Deme                1         2         3
Allele
 101                6        11         7
 121                8         9         0
 132                0         0         7
 136                6         0         6
               -------   -------   -------
                  20        20        20

-----------------------------------
Relative_frequency_distribution
-----------------------------------
Deme                1         2         3
Allele
 101            0.300     0.550     0.350
 121            0.400     0.450     0.000
 132            0.000     0.000     0.350
 136            0.300     0.000     0.300

-----------------------------------
Measures_of_variation_within_demes
-----------------------------------

DIVERSITY_v_2
Deme                1         2         3
               2.941     1.980     2.985

TOTAL_POPULATION_DIFFERENTIATION_delta_T
Deme                1         2         3
               0.695     0.521     0.700
EVENNESS_e_FOR_FINITE_POPULATION_SIZE
Deme                1         2         3
absolute_e      0.950     0.950     1.000
relative_e      0.900     0.900     1.000
for_No.types        3         2         3
EVENNESS_e_FOR_INFINITE_POPULATION_SIZE
Deme                1         2         3
absolute_e      0.933     0.950     0.967
```

```
relative_e     0.867     0.900     0.933
for_No.types       3         2         3

-----------------------------------
Measures_of_variation_between_demes
-----------------------------------

GENETIC_DISTANCE_d_0
Deme               1         2         3
          ------------------------------
         1 |   0.000
         2 |   0.300     0.000
         3 |   0.400     0.650     0.000

SUBPOPULATION_DIFFERENTIATION_D_j,_delta
- RELATIVE_SUBPOPULATION_SIZE_PROPORTIONAL_TO_DEME_SIZE
Deme               1         2         3
       Cj=    0.333     0.333     0.333
       Dj=    0.325     0.475     0.500

        delta=    0.433

SUBPOPULATION_DIFFERENTIATION_D_j,_delta
- RELATIVE_SUBPOPULATION_SIZES_ALL_EQUAL_TO_(1/NO._SUBPOPS)
Deme               1         2         3
       Cj=    0.333     0.333     0.333
       Dj=    0.325     0.475     0.500

        delta=    0.433

TEST_OF_HOMOGENEITY
Deme               1         2         3
Allele/   Sum ------------------------------
   101     24 O    6        11         7
              E    8.00      8.00      8.00
   121     17 O    8         9         0
              E    5.67      5.67      5.67
   132      7 O    0         0         7
              E    2.33      2.33      2.33
   136     12 O    6         0         6
              E    4.00      4.00      4.00
         -----------------------------------
            60 |   20        20        20

                                 Level_of   C.V._of_CHI**2
Test_statistics                 significance   (DF=   6)
-------------------------------------------------------------
G          =   40.641 ***       0.050(*  )      12.592
X**2       =   30.338 ***       0.010(** )      16.812
                                0.001(***)      22.458
*****<_WARNING_>*****<_WARNING_>*****<_WARNING_>*****<_WARNING_>*****
** Test_statistics_are_inflated_due_to_expected_frequencies_less   **
** than  5 and_may_falsely_recommend_rejection_of_hypothesis.      **
** Suggestion:_Pool_alleles_in_input_data_and_recalculate.        **
*****<_WARNING_>*****<_WARNING_>*****<_WARNING_>*****<_WARNING_>*****

=========================================================================
Genotype_frequencies
=========================================================================
Deme               1         2         3
Gam.sex.spec.?    no        no        no
```

```
Deme_size          10         10         10
No._identified     10         10         10
alpha           0.334      0.334      0.334
No._unknown         0          0          0

-----------------------------------
Absolute_frequency_distribution
-----------------------------------
Deme                1          2          3
Genotype
 101 x 101          1          3          2
 101 x 121          2          5          0
 101 x 136          2          0          3
 121 x 121          2          2          0
 121 x 136          2          0          0
 132 x 132          0          0          2
 132 x 136          0          0          3
 136 x 136          1          0          0
               -------    -------    -------
                   10         10         10

-----------------------------------
Relative_frequency_distribution
-----------------------------------
Deme                1          2          3
Genotype
 101 x 101      0.100      0.300      0.200
 101 x 121      0.200      0.500      0.000
 101 x 136      0.200      0.000      0.300
 121 x 121      0.200      0.200      0.000
 121 x 136      0.200      0.000      0.000
 132 x 132      0.000      0.000      0.200
 132 x 136      0.000      0.000      0.300
 136 x 136      0.100      0.000      0.000

-----------------------------------
Measures_of_variation_within_demes
-----------------------------------

DIVERSITY_v_2
Deme                1          2          3
               5.556      2.632      3.846

TOTAL_POPULATION_DIFFERENTIATION_delta_T
Deme                1          2          3
               0.911      0.689      0.822
EVENNESS_e_FOR_FINITE_POPULATION_SIZE
Deme                1          2          3
absolute_e      1.000      0.900      1.000
relative_e      1.000      0.800      1.000
for_No.types        6          3          4
EVENNESS_e_FOR_INFINITE_POPULATION_SIZE
Deme                1          2          3
absolute_e      0.900      0.833      0.900
relative_e      0.800      0.667      0.800
for_No.types        5          3          4

-----------------------------------
Measures_of_variation_between_demes
-----------------------------------
```

```
GENETIC_DISTANCE_d_0
Deme              1         2         3
          ------------------------------
         1 |   0.000
         2 |   0.500     0.000
         3 |   0.700     0.800     0.000

SUBPOPULATION_DIFFERENTIATION_D_j,_delta
- RELATIVE_SUBPOPULATION_SIZE_PROPORTIONAL_TO_DEME_SIZE
Deme              1         2         3
       Cj=    0.333     0.333     0.333
       Dj=    0.450     0.650     0.700

        delta=    0.600

SUBPOPULATION_DIFFERENTIATION_D_j,_delta
- RELATIVE_SUBPOPULATION_SIZES_ALL_EQUAL_TO_(1/NO._SUBPOPS)
Deme              1         2         3
       Cj=    0.333     0.333     0.333
       Dj=    0.450     0.650     0.700

        delta=    0.600

TEST_OF_HOMOGENEITY
Deme              1         2         3
Genotype/ Sum ------------------------------
101 x101    6 O    1         3         2
              E    2.00      2.00      2.00
101 x121    7 O    2         5         0
              E    2.33      2.33      2.33
101 x136    5 O    2         0         3
              E    1.67      1.67      1.67
121 x121    4 O    2         2         0
              E    1.33      1.33      1.33
121 x136    2 O    2         0         0
              E    0.67      0.67      0.67
132 x132    2 O    0         0         2
              E    0.67      0.67      0.67
132 x136    3 O    0         0         3
              E    1.00      1.00      1.00
136 x136    1 O    1         0         0
              E    0.33      0.33      0.33
        -----------------------------------
           30 |   10        10        10

                                  Level_of   C.V._of_CHI**2
Test_statistics                 significance   (DF=  14)
-----------------------------------------------------------
G          =   33.129 **         0.050(*  )      23.685
X**2       =   27.229 *          0.010(** )      29.141
                                 0.001(***)      36.123
*****<_WARNING_>*****<_WARNING_>*****<_WARNING_>*****<_WARNING_>*****
** Test_statistics_are_inflated_due_to_expected_frequencies_less   **
** than  5 and_may_falsely_recommend_rejection_of_hypothesis.      **
** Suggestion:_Pool_alleles_in_input_data_and_recalculate.        **
*****<_WARNING_>*****<_WARNING_>*****<_WARNING_>*****<_WARNING_>*****

=========================================================================
Heterozygosity
=========================================================================
Deme              1         2         3
```

```
Deme size          10         10         10
No. identified     10         10         10
No. unknown         0          0          0

PROPORTION OF HETEROZYGOSITY
Deme                1          2          3
                0.600      0.500      0.600

CONDITIONAL HETEROZYGOSITY
Deme                1          2          3
                0.600      0.556      0.600

**************************************************************************
Locus   1 - Locus1
Deme    1 - Pop1
**************************************************************************
Deme_size                                  10
No._individuals_of_identified_genotype     10
No._individuals_of_unknown_genotype         0

==========================================================================
Test_of_Hardy-Weinberg_Structure
==========================================================================

Allele    101       121       136     Freq.
       -------------------------------------
  101  O    1         2         2    |    6
       E    0.90      2.40      1.80 |
  121  O              2         2    |    8
       E              1.60      2.40 |
  136  O                        1    |    6
       E                        0.90 |
       -------------------------------------
                                     |   20

                                 Level_of   C.V._of CHI**2
Test_statistics                 significance   (DF=   3)
-----------------------------------------------------------
G          =    0.277 n.s.      0.050(*  )       7.815
X**2       =    0.278 n.s.      0.010(** )      11.345
                                0.001(***)      16.266

==========================================================================
Test_of_Hardy-Weinberg_Heterozygosity
==========================================================================

      Homozygotes Heterozygotes
      -------------------------
       O    4         6
       E    3.40      6.60
      -------------------------
      Homozygote_excess_over_Hardy-Weinberg_expectation

                                 Level_of   C.V._of_CHI**2
Test_statistics                 significance   (DF=   1)
-----------------------------------------------------------
G          =    0.156 n.s.      0.050(*  )       3.841
X**2       =    0.160 n.s.      0.010(** )       6.635
X**2(c=.5) =    0.004 n.s.      0.001(***)      10.828
```

G Output file `example.txt-tab.txt`

Beginning of the output file for single-locus genotypes at Locus 1 in the sample input file `example.txt-out.txt`

```
GSED_INPUT__FILE "path/example.txt"
GSED_OUTPUT_FILE "path/example.txt-out.txt"
Date 12-Apr-2010 12:24:42

LOCUS_NO. 1     "Locus1"
ALLELE_FREQUENCIES

Absolute_frequencies:
Deme_No.                1              2              3
Type_1                  6             11              7           "{ 101 }"
Type_2                  8              9              0           "{ 121 }"
Type_3                  0              0              7           "{ 132 }"
Type_4                  6              0              6           "{ 136 }"
Sum                    20             20             20

Relative_frequencies:
Deme_No.                1              2              3
Type_1              0.300          0.550          0.350           "{ 101 }"
Type_2              0.400          0.450          0.000           "{ 121 }"
Type_3              0.000          0.000          0.350           "{ 132 }"
Type_4              0.300          0.000          0.300           "{ 136 }"

Measures_of_genetic_variation:
Deme_No.    DemeName                DemeSize      Alpha  Alpha-HWP  v2-Divers
1          "Pop1                "          20     0.2010     0.1170     2.9412
2          "Pop2                "          20     0.2010     0.1170     1.9802
3          "Pop3                "          20     0.2010     0.1170     2.9851
```

– *wrapped* –

```
     deltaT  EvnFinAbs  EvnFinRel  EvnFinNum  EvnInfAbs  EvnInfRel  EvnInfNum
     0.6947     0.9500     0.9000     3.0000     0.9333     0.8667     3.0000
     0.5211     0.9500     0.9000     2.0000     0.9500     0.9000     2.0000
     0.7000     1.0000     1.0000     3.0000     0.9667     0.9333     3.0000
```

– *wrapped* –

```
   CjDemSiz   DjDemSiz   deltaDmS   CjEquSiz   DjEquSiz   deltaEqS
     0.3333     0.3250     0.4333     0.3333     0.3250     0.4333
     0.3333     0.4750     0.4333     0.3333     0.4750     0.4333
     0.3333     0.5000     0.4333     0.3333     0.5000     0.4333

Genetic_distance_d_0
Deme_No.                1              2              3
1                   0.0000
2                   0.3000         0.0000
3                   0.4000         0.6500         0.0000

LOCUS_NO. 1    "Locus1"
GENOTYPE_FREQUENCIES

Absolute_frequency_distribution:
Deme_No.        1       2       3
Type_1       1       3       2      "{ 101 101  }"
Type_2       2       5       0      "{ 101 121  }"
Type_3       2       0       3      "{ 101 136  }"
```

```
Type_4       2       2       0       "{ 121 121  }"
Type_5       2       0       0       "{ 121 136  }"
Type_6       0       0       2       "{ 132 132  }"
Type_7       0       0       3       "{ 132 136  }"
Type_8       1       0       0       "{ 136 136  }"
Sum          10      10      10

Relative_frequency_distribution:
Deme_No.         1       2       3
Type_1       0.100       0.300       0.200       "{ 101 101  }"
Type_2       0.200       0.500       0.000       "{ 101 121  }"
Type_3       0.200       0.000       0.300       "{ 101 136  }"
Type_4       0.200       0.200       0.000       "{ 121 121  }"
Type_5       0.200       0.000       0.000       "{ 121 136  }"
Type_6       0.000       0.000       0.200       "{ 132 132  }"
Type_7       0.000       0.000       0.300       "{ 132 136  }"
Type_8       0.100       0.000       0.000       "{ 136 136  }"

Measures_of_genetic_variation:
Deme_No.    DemeName             DemeSize      Alpha  Alpha-HWP  v2-Divers
1           "Pop1                "      10     0.3340     0.2010     5.5556
2           "Pop2                "      10     0.3340     0.2010     2.6316
3           "Pop3                "      10     0.3340     0.2010     3.8462
```

– *wrapped* –

```
  deltaT  EvnFinAbs  EvnFinRel  EvnFinNum  EvnInfAbs  EvnInfRel  EvnInfNum
  0.9111     1.0000     1.0000     6.0000     0.9000     0.8000     5.0000
  0.6889     0.9000     0.8000     3.0000     0.8333     0.6667     3.0000
  0.8222     1.0000     1.0000     4.0000     0.9000     0.8000     4.0000
```

– *wrapped* –

```
CjDemSiz   DjDemSiz   deltaDmS   CjEquSiz   DjEquSiz   deltaEqS  PropHeter
  0.3333     0.4500     0.6000     0.3333     0.4500     0.6000     0.6000
  0.3333     0.6500     0.6000     0.3333     0.6500     0.6000     0.5000
  0.3333     0.7000     0.6000     0.3333     0.7000     0.6000     0.6000
```

– *wrapped* –

```
CondHeter
   0.6000
   0.5556
   0.6000

Genetic_distance_d_0
Deme_No.      1          2          3
1        0.0000
2        0.5000     0.0000
3        0.7000     0.8000     0.0000
```

H Output file `example.txt-multi-out.txt`

The output file named `example.txt-multi-out.txt` for multilocus genotypes in the sample input file `example.txt-out.txt`

```
**********************************************************************
*       GGGGGGG  SSSSSSS  EEEEEEE  DDDDDD          Genetic            *
*       G     G  S        E        D     D         Structures         *
*       G        S        E        D      D        from               *
*       G        SSSSSSS  EEEE     D      D        Electrophoresis    *
*       G  GGGG         S  E        D      D        Data               *
*       G     G         S  E        D     D         Version 3.0beta    *
*       GGGGGGG  SSSSSSS  EEEEEEE  DDDDDD          April 2010         *
**********************************************************************
GSED Copyright 1990-2010 Elizabeth M. Gillet, egillet@gwdg.de
     <www.uni-goettingen.de/de/95607.html>
GSED is free software: you can redistribute it under the terms of the
GNU General Public License (GPL) v.3, as published by the Free Software
Foundation. GSED is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of merchant-
ability or fitness for a particular purpose. Reassembling is not permit-
ted. See the GNU General Public License (GPL). A copy of the GNU GPL is
contained in the file COPYING, or see <www.gnu.org/licenses>.
Plot and widget routines: DISLIN <www.dislin.de>
Compiler: GNU Fortran (SUSE Linux) 4.3.2 gcc-4_3-branch revision 141291]
----------------------------------------------------------------------
Input_file: path/example.txt
Date 19-Apr-2010 15:52:46

Deme    1 : Pop1
Deme    2 : Pop2
Deme    3 : Pop3

Combination_No.   1 : Locus   1 - Locus1
                      Locus   2 - Locus2
                      Locus   3 - Locus3
                      Locus   4 - Locus4
----------------------------------------------------------------------
Abbreviations
----------------------------------------------------------------------
O_or_E           Observed_or_Expected_absolute_frequency_in_a_test
{_}              Denotes_multilocus_haplotype_or_genotype
NA               Denotes_undefinable_parameter_value
Gam.sex.spec.?   Abbreviation_of_"Gametic_sex_specification?"
               = "yes",_if_maternal/paternal_alleles_distinguishable
               = "no"_,_otherwise.
alpha            All_alleles/haplotypes/genotypes_of_relative_frequency
                 not_less_than_"alpha"_in_deme_appear_in_sample
                 (with replacement) with_probability>=0.95
alpha-HWP        As_above,_if_genotypes_in_deme_are_in
                 Hardy-Weinberg-Proportions_(HWP)

**********************************************************************
Combination_No.  1 :      Locus   1 - Locus1
                          Locus   2 - Locus2
                          Locus   3 - Locus3
                          Locus   4 - Locus4
**********************************************************************
** Haplotype_frequencies_not_specifiable

======================================================================
```

```
Genotype_frequencies
============================================================================
Deme                1          2          3
Gam.sex.spec.?     no         no         no
Deme_size          10         10         10
No._identified     10         10          9
alpha           0.334      0.334      0.334
No._unknown         0          0          1

-----------------------------------
Absolute_frequency_distribution
-----------------------------------
Deme                1          2          3
Genotype
   1. {101 101   76  76  121 193   36  48   }
                    0          1          0
   2. {101 101   76  83   89 106   36  48   }
                    0          1          0
   3. {101 101   76  83  106 106   42  42   }
                    0          0          1
   4. {101 101   76 102  106 106   48  48   }
                    0          0          1
   5. {101 101   83  83  106 106   36  48   }
                    0          1          0
   6. {101 101   83 102  106 193   36  36   }
                    1          0          0
   7. {101 121   76  95   89  89   36  48   }
                    0          1          0
   8. {101 121   76  95  118 121   42  42   }
                    1          0          0
   9. {101 121   76 102   89 193   36  48   }
                    0          1          0
  10. {101 121   76 102  121 121   36  48   }
                    0          1          0
  11. {101 121   83  95  106 193   36  48   }
                    0          1          0
  12. {101 121   83 102   89 193   36  48   }
                    0          1          0
  13. {101 121   95  95   89 106   42  48   }
                    1          0          0
  14. {101 136   83  83  106 193   42  48   }
                    1          0          0
  15. {101 136   83  95  121 121   36  42   }
                    0          0          1
  16. {101 136   83 102   89  89   36  42   }
                    0          0          1
  17. {101 136   95 102  106 121   42  48   }
                    1          0          0
  18. {121 121   76  76  121 193   36  42   }
                    0          1          0
  19. {121 121   76  83   89 193   36  36   }
                    1          0          0
  20. {121 121   83  83   89  89   36  42   }
                    0          1          0
  21. {121 121   83  95  118 193   42  42   }
                    1          0          0
  22. {121 136   83 102  106 118   36  42   }
                    1          0          0
  23. {121 136   83 102  106 193   36  36   }
                    1          0          0
  24. {132 132   76  76   89  89   42  42   }
```

```
                    0          0          1
  25. {132 132   95  95   89 121   42  48   }
                    0          0          1
  26. {132 136   76  83   89 106   36  42   }
                    0          0          1
  27. {132 136   83  83   89 106   48  48   }
                    0          0          1
  28. {132 136   95  95  121 121   36  48   }
                    0          0          1
  29. {136 136   76  95  118 121   42  48   }
                    1          0          0
                -------    -------    -------
                   10         10          9

-----------------------------------
Relative_frequency_distribution
-----------------------------------
Deme                  1          2          3
Genotype
   1. {101 101   76  76  121 193   36  48   }
                0.000      0.100      0.000
   2. {101 101   76  83   89 106   36  48   }
                0.000      0.100      0.000
   3. {101 101   76  83  106 106   42  42   }
                0.000      0.000      0.111
   4. {101 101   76 102  106 106   48  48   }
                0.000      0.000      0.111
   5. {101 101   83  83  106 106   36  48   }
                0.000      0.100      0.000
   6. {101 101   83 102  106 193   36  36   }
                0.100      0.000      0.000
   7. {101 121   76  95   89  89   36  48   }
                0.000      0.100      0.000
   8. {101 121   76  95  118 121   42  42   }
                0.100      0.000      0.000
   9. {101 121   76 102   89 193   36  48   }
                0.000      0.100      0.000
  10. {101 121   76 102  121 121   36  48   }
                0.000      0.100      0.000
  11. {101 121   83  95  106 193   36  48   }
                0.000      0.100      0.000
  12. {101 121   83 102   89 193   36  48   }
                0.000      0.100      0.000
  13. {101 121   95  95   89 106   42  48   }
                0.100      0.000      0.000
  14. {101 136   83  83  106 193   42  48   }
                0.100      0.000      0.000
  15. {101 136   83  95  121 121   36  42   }
                0.000      0.000      0.111
  16. {101 136   83 102   89  89   36  42   }
                0.000      0.000      0.111
  17. {101 136   95 102  106 121   42  48   }
                0.100      0.000      0.000
  18. {121 121   76  76  121 193   36  42   }
                0.000      0.100      0.000
  19. {121 121   76  83   89 193   36  36   }
                0.100      0.000      0.000
  20. {121 121   83  83   89  89   36  42   }
                0.000      0.100      0.000
  21. {121 121   83  95  118 193   42  42   }
                0.100      0.000      0.000
```

```
  22. {121 136   83 102  106 118   36  42   }
               0.100     0.000     0.000
  23. {121 136   83 102  106 193   36  36   }
               0.100     0.000     0.000
  24. {132 132   76  76   89  89   42  42   }
               0.000     0.000     0.111
  25. {132 132   95  95   89 121   42  48   }
               0.000     0.000     0.111
  26. {132 136   76  83   89 106   36  42   }
               0.000     0.000     0.111
  27. {132 136   83  83   89 106   48  48   }
               0.000     0.000     0.111
  28. {132 136   95  95  121 121   36  48   }
               0.000     0.000     0.111
  29. {136 136   76  95  118 121   42  48   }
               0.100     0.000     0.000

-----------------------------------
Measures_of_variation_within_demes
-----------------------------------

DIVERSITY_v_2
Deme               1         2         3
              10.000    10.000     9.000

TOTAL_POPULATION_DIFFERENTIATION_delta_T
Deme               1         2         3
               1.000     1.000     1.000
EVENNESS_e_FOR_FINITE_POPULATION_SIZE
Deme               1         2         3
absolute_e     1.000     1.000     1.000
relative_e     1.000     1.000     1.000
for_No.types      10        10         9
EVENNESS_e_FOR_INFINITE_POPULATION_SIZE
Deme               1         2         3
absolute_e     1.000     1.000     1.000
relative_e     1.000     1.000     1.000
for_No.types      10        10         9

-----------------------------------
Measures_of_variation_between_demes
-----------------------------------

GENETIC_DISTANCE_d_0
Deme               1         2         3
          ------------------------------
         1 |   0.000
         2 |   1.000     0.000
         3 |   1.000     1.000     0.000

SUBPOPULATION_DIFFERENTIATION_D_j,_delta
- RELATIVE_SUBPOPULATION_SIZE_PROPORTIONAL_TO_DEME_SIZE
Deme               1         2         3
       Cj=    0.345     0.345     0.310
       Dj=    1.000     1.000     1.000

        delta=    1.000

SUBPOPULATION_DIFFERENTIATION_D_j,_delta
- RELATIVE_SUBPOPULATION_SIZES_ALL_EQUAL_TO_(1/NO._SUBPOPS)
Deme               1         2         3
```

```
       Cj=    0.333     0.333     0.333
       Dj=    1.000     1.000     1.000

        delta=    1.000

TEST_OF_HOMOGENEITY
Deme                1         2         3
Genotype/ Sum -------------------------------
    1.  {101 101   76  76  121 193   36  48   }
           1 O     0         1         0
             E     0.34      0.34      0.31
    2.  {101 101   76  83   89 106   36  48   }
           1 O     0         1         0
             E     0.34      0.34      0.31
    3.  {101 101   76  83  106 106   42  42   }
           1 O     0         0         1
             E     0.34      0.34      0.31
    4.  {101 101   76 102  106 106   48  48   }
           1 O     0         0         1
             E     0.34      0.34      0.31
    5.  {101 101   83  83  106 106   36  48   }
           1 O     0         1         0
             E     0.34      0.34      0.31
    6.  {101 101   83 102  106 193   36  36   }
           1 O     1         0         0
             E     0.34      0.34      0.31
    7.  {101 121   76  95   89  89   36  48   }
           1 O     0         1         0
             E     0.34      0.34      0.31
    8.  {101 121   76  95  118 121   42  42   }
           1 O     1         0         0
             E     0.34      0.34      0.31
    9.  {101 121   76 102   89 193   36  48   }
           1 O     0         1         0
             E     0.34      0.34      0.31
   10.  {101 121   76 102  121 121   36  48   }
           1 O     0         1         0
             E     0.34      0.34      0.31
   11.  {101 121   83  95  106 193   36  48   }
           1 O     0         1         0
             E     0.34      0.34      0.31
   12.  {101 121   83 102   89 193   36  48   }
           1 O     0         1         0
             E     0.34      0.34      0.31
   13.  {101 121   95  95   89 106   42  48   }
           1 O     1         0         0
             E     0.34      0.34      0.31
   14.  {101 136   83  83  106 193   42  48   }
           1 O     1         0         0
             E     0.34      0.34      0.31
   15.  {101 136   83  95  121 121   36  42   }
           1 O     0         0         1
             E     0.34      0.34      0.31
   16.  {101 136   83 102   89  89   36  42   }
           1 O     0         0         1
             E     0.34      0.34      0.31
   17.  {101 136   95 102  106 121   42  48   }
           1 O     1         0         0
             E     0.34      0.34      0.31
   18.  {121 121   76  76  121 193   36  42   }
           1 O     0         1         0
```

```
            E    0.34      0.34      0.31
   19.  {121 121   76  83   89 193   36  36   }
          1 O    1         0         0
            E    0.34      0.34      0.31
   20.  {121 121   83  83   89  89   36  42   }
          1 O    0         1         0
            E    0.34      0.34      0.31
   21.  {121 121   83  95  118 193   42  42   }
          1 O    1         0         0
            E    0.34      0.34      0.31
   22.  {121 136   83 102  106 118   36  42   }
          1 O    1         0         0
            E    0.34      0.34      0.31
   23.  {121 136   83 102  106 193   36  36   }
          1 O    1         0         0
            E    0.34      0.34      0.31
   24.  {132 132   76  76   89  89   42  42   }
          1 O    0         0         1
            E    0.34      0.34      0.31
   25.  {132 132   95  95   89 121   42  48   }
          1 O    0         0         1
            E    0.34      0.34      0.31
   26.  {132 136   76  83   89 106   36  42   }
          1 O    0         0         1
            E    0.34      0.34      0.31
   27.  {132 136   83  83   89 106   48  48   }
          1 O    0         0         1
            E    0.34      0.34      0.31
   28.  {132 136   95  95  121 121   36  48   }
          1 O    0         0         1
            E    0.34      0.34      0.31
   29.  {136 136   76  95  118 121   42  48   }
          1 O    1         0         0
            E    0.34      0.34      0.31
        ------------------------------------
           29 |   10        10         9

                                    Level_of   C.V._of_CHI**2
Test_statistics                   significance   (DF=  56)
-------------------------------------------------------------
G          =   63.650 n.s.       0.050(*  )      74.468
X**2       =   58.000 n.s.       0.010(** )      83.513
                                 0.001(***)      94.461

=============================================================================
Heterozygosity
=============================================================================
Deme               1         2         3
Deme size         10        10        10
No. identified    10        10         9
No. unknown        0         0         1

DISTRIBUTION OF DEGREE OF HETEROZYGOSITY
OF PROBES WITH COMPLETELY IDENTIFIED  4-LOCUS GENOTYPES
Deme               1         2         3
hloc=   0          0         0         1
           (   0.000)(   0.000)(   0.111)
hloc=   1          0         2         2
           (   0.000)(   0.200)(   0.222)
hloc=   2          3         2         3
           (   0.300)(   0.200)(   0.333)
```

```
hloc=   3          5         3         2
           (   0.500)(   0.300)(   0.222)
hloc=   4          2         3         1
           (   0.200)(   0.300)(   0.111)
             --------  --------  --------
                 10        10         9

AVERAGE DEGREE OF HETEROZYGOSITY
OF PROBES WITH COMPLETELY IDENTIFIED  4-LOCUS GENOTYPES
Deme               1         2         3
              0.725     0.675     0.500

-----------------------------------------------------------
** Combination  1 is_not_a_single_locus:
** Test_of_Hardy-Weinberg_structure_is_not_possible
** Test_of_product_structure_is_not_possible
-----------------------------------------------------------
```

I Output file `example.txt-multi-tab.txt`

Example of output file named `example.txt-multi-tab.txt` for the multilocus genotype frequencies at all loci in the sample input file `example.txt`. The long lines under `Measures_of_genetic_variation:` are wrapped in order to fit the table on the page.

```
GSED_INPUT__FILE "path/example.txt"
GSED_OUTPUT_FILE "path/example.txt-multi-out.txt"
Date 19-Apr-2010 15:52:46

MULTILOCUS_COMBINATION_OF_LOCI:
LOCUS_NO. 1        "Locus1"
LOCUS_NO. 2        "Locus2"
LOCUS_NO. 3        "Locus3"
LOCUS_NO. 4        "Locus4"
GENOTYPE_FREQUENCIES

Absolute_frequency_distribution:
Deme_No.      1       2       3
Type_1        0       1       0    "{ 101 101   76  76  121 193   36  48  }"
Type_2        0       1       0    "{ 101 101   76  83   89 106   36  48  }"
Type_3        0       0       1    "{ 101 101   76  83  106 106   42  42  }"
Type_4        0       0       1    "{ 101 101   76 102  106 106   48  48  }"
Type_5        0       1       0    "{ 101 101   83  83  106 106   36  48  }"
Type_6        1       0       0    "{ 101 101   83 102  106 193   36  36  }"
Type_7        0       1       0    "{ 101 121   76  95   89  89   36  48  }"
Type_8        1       0       0    "{ 101 121   76  95  118 121   42  42  }"
Type_9        0       1       0    "{ 101 121   76 102   89 193   36  48  }"
Type_10       0       1       0    "{ 101 121   76 102  121 121   36  48  }"
Type_11       0       1       0    "{ 101 121   83  95  106 193   36  48  }"
Type_12       0       1       0    "{ 101 121   83 102   89 193   36  48  }"
Type_13       1       0       0    "{ 101 121   95  95   89 106   42  48  }"
Type_14       1       0       0    "{ 101 136   83  83  106 193   42  48  }"
Type_15       0       0       1    "{ 101 136   83  95  121 121   36  42  }"
Type_16       0       0       1    "{ 101 136   83 102   89  89   36  42  }"
Type_17       1       0       0    "{ 101 136   95 102  106 121   42  48  }"
Type_18       0       1       0    "{ 121 121   76  76  121 193   36  42  }"
Type_19       1       0       0    "{ 121 121   76  83   89 193   36  36  }"
Type_20       0       1       0    "{ 121 121   83  83   89  89   36  42  }"
Type_21       1       0       0    "{ 121 121   83  95  118 193   42  42  }"
Type_22       1       0       0    "{ 121 136   83 102  106 118   36  42  }"
Type_23       1       0       0    "{ 121 136   83 102  106 193   36  36  }"
Type_24       0       0       1    "{ 132 132   76  76   89  89   42  42  }"
Type_25       0       0       1    "{ 132 132   95  95   89 121   42  48  }"
Type_26       0       0       1    "{ 132 136   76  83   89 106   36  42  }"
Type_27       0       0       1    "{ 132 136   83  83   89 106   48  48  }"
Type_28       0       0       1    "{ 132 136   95  95  121 121   36  48  }"
Type_29       1       0       0    "{ 136 136   76  95  118 121   42  48  }"
Sum          10      10       9

Relative_frequency_distribution:
Deme_No.      1       2       3
Type_1    0.000   0.100   0.000    "{ 101 101   76  76  121 193   36  48  }"
Type_2    0.000   0.100   0.000    "{ 101 101   76  83   89 106   36  48  }"
Type_3    0.000   0.000   0.111    "{ 101 101   76  83  106 106   42  42  }"
Type_4    0.000   0.000   0.111    "{ 101 101   76 102  106 106   48  48  }"
```

```
Type_5    0.000   0.100   0.000    "{ 101 101   83  83  106 106   36  48  }"
Type_6    0.100   0.000   0.000    "{ 101 101   83 102  106 193   36  36  }"
Type_7    0.000   0.100   0.000    "{ 101 121   76  95   89  89   36  48  }"
Type_8    0.100   0.000   0.000    "{ 101 121   76  95  118 121   42  42  }"
Type_9    0.000   0.100   0.000    "{ 101 121   76 102   89 193   36  48  }"
Type_10   0.000   0.100   0.000    "{ 101 121   76 102  121 121   36  48  }"
Type_11   0.000   0.100   0.000    "{ 101 121   83  95  106 193   36  48  }"
Type_12   0.000   0.100   0.000    "{ 101 121   83 102   89 193   36  48  }"
Type_13   0.100   0.000   0.000    "{ 101 121   95  95   89 106   42  48  }"
Type_14   0.100   0.000   0.000    "{ 101 136   83  83  106 193   42  48  }"
Type_15   0.000   0.000   0.111    "{ 101 136   83  95  121 121   36  42  }"
Type_16   0.000   0.000   0.111    "{ 101 136   83 102   89  89   36  42  }"
Type_17   0.100   0.000   0.000    "{ 101 136   95 102  106 121   42  48  }"
Type_18   0.000   0.100   0.000    "{ 121 121   76  76  121 193   36  42  }"
Type_19   0.100   0.000   0.000    "{ 121 121   76  83   89 193   36  36  }"
Type_20   0.000   0.100   0.000    "{ 121 121   83  83   89  89   36  42  }"
Type_21   0.100   0.000   0.000    "{ 121 121   83  95  118 193   42  42  }"
Type_22   0.100   0.000   0.000    "{ 121 136   83 102  106 118   36  42  }"
Type_23   0.100   0.000   0.000    "{ 121 136   83 102  106 193   36  36  }"
Type_24   0.000   0.000   0.111    "{ 132 132   76  76   89  89   42  42  }"
Type_25   0.000   0.000   0.111    "{ 132 132   95  95   89 121   42  48  }"
Type_26   0.000   0.000   0.111    "{ 132 136   76  83   89 106   36  42  }"
Type_27   0.000   0.000   0.111    "{ 132 136   83  83   89 106   48  48  }"
Type_28   0.000   0.000   0.111    "{ 132 136   95  95  121 121   36  48  }"
Type_29   0.100   0.000   0.000    "{ 136 136   76  95  118 121   42  48  }"

Measures_of_genetic_variation:
Deme_No.     DemeName                  DemeSize       Alpha   v2-Divers
1           "Pop1                "          10      0.3340     10.0000
2           "Pop2                "          10      0.3340     10.0000
3           "Pop3                "           9      0.3340      9.0000
```

– *wrapped* –

```
     deltaT  EvnFinAbs  EvnFinRel  EvnFinNum  EvnInfAbs  EvnInfRel  EvnInfNum
     1.0000     1.0000     1.0000    10.0000     1.0000     1.0000    10.0000
     1.0000     1.0000     1.0000    10.0000     1.0000     1.0000    10.0000
     1.0000     1.0000     1.0000     9.0000     1.0000     1.0000     9.0000
```

– *wrapped* –

```
   CjDemSiz   DjDemSiz   deltaDmS   CjEquSiz   DjEquSiz   deltaEqS  MeanHeter
     0.3448     1.0000     1.0000     0.3333     1.0000     1.0000     0.7250
     0.3448     1.0000     1.0000     0.3333     1.0000     1.0000     0.6750
     0.3103     1.0000     1.0000     0.3333     1.0000     1.0000     0.5000

Genetic_distance_d_0
Deme_No.            1          2          3
1              0.0000
2              1.0000     0.0000
3              1.0000     1.0000     0.0000
```