User's Guide

ENDOG v4.8

A Computer Program for Monitoring Genetic Variability of Populations Using Pedigree Information

Latest update of this guide on November 10, 2010

Juan Pablo Gutiérrez, DVM, PhD[1]

Departamento de Producción Animal.
Facultad de Veterinaria.
Universidad Complutense de Madrid
Avda. Puerta de Hierro s/n
E-28040-Madrid, Spain.
Phone and Fax: +34 913943767.
E-mail: gutgar@vet.ucm.es

Félix Goyache, DVM, PhD

Área de Genética y Reproducción Animal
SERIDA-Somió
C/ Camino de los Claveles 604
E-33203 Gijón (Asturias), Spain.
Phone: +34 985195303. Fax: +34 985195310
E-mail: fgoyache@serida.org

Isabel Cervantes, DVM, PhD

Departamento de Producción Animal.
Facultad de Veterinaria.
Universidad Complutense de Madrid
Avda. Puerta de Hierro s/n
E-28040-Madrid, Spain.
Phone and Fax: +34 913943773.
E-mail: icervantes@vet.ucm.es

[1] To whom correspondence should be directed

INDEX

1.- Introduction

1.1 Purpose and General Comments

The knowledge of genetic variability within populations has received increasing attention over recent years (Wooliams et al., 2002). In populations under selective pressure, the inbreeding within the progeny of reproducing individuals can be higher than that expected under pure genetic drift. On the other hand, the goal in conservation programmes for endangered breeds is to restrain the rate of inbreeding. Considering both selection and conservation, some simple demographic parameters have a large impact on the evolution of the genetic variability and largely depend on the management of the population (Gutiérrez et al., 2003; Goyache et al., 2003). The computation of effective population size (N_e; Falconer and Mackay, 1996) is a key parameter for describing genetic diversity in animal populations but also for predictive purposes. In addition, we can ascertain the extent to which an inappropriate mating policy leads to structuring of populations (Caballero and Toro, 2002). When the ENDOG project started, few computer routines were available to test the evolution of the genetic variability of populations using pedigree information (Boichard, 2002). ENDOG (current version 4.8) is a population genetics computer program that conducts several demographic and genetic analyses on pedigree information in a user friendly environment. The program will help researchers or those responsible for management of populations to monitor the changes in genetic variability and population structure with a limited amount of prior preparation of datasets. Although written primarily as a population monitoring package, ENDOG does offer a number of features that may be of interest to teachers and students to develop an in-depth understanding of the important statistical concepts and procedures relevant to population genetic analysis

ENDOG is tributary of a suite of FORTRAN 77 routines which were widely distributed and used among Spanish groups (Gutiérrez et al., 2003). ENDOG (version 2.0) was used for the calculations described in Goyache et al. (2003). Version 3.2 of the program included the possibility of computing F-statistics (Wright, 1978) and the Individual menu. From version 4.2, features such as the computation of recent inbreeding, the assessment of the completeness of the pedigree and the estimation of the effective size of the analysed population using the family variances were included in the program. ENDOG 4.6 included a number of improvements, namely affecting the computation of N_e using regression approaches and the formula by Hill (1979) on family variances, but especially the computation of individual increase of inbreeding (ΔF_i) and partial inbreeding coefficients (Lacy et al., 1996). The present version of ENDOG (v4.8) includes the imporved in the estimation of the 'realised' N_e proposed by Gutiérrez et al. (2009) from individual increase of inbreeding, the computation of N_e from paired increase in coancestry (Cervantes et al., 2010), the assesment of the contributions of predefined subpopulations to total diversity (Caballero and Toro, 2002) and a methodology for adjustment of within- and betwen-subpopulations coancestries by subpopulations size (Bartolomé et al., 2010). ENDOG has been written in VisualBasic^{TM} language and runs under Windows 95/98/2000/NT/XP versions. ENDOG has been tested under Windows Seven. The program works appropriately. A setup menu will guide users when installing the program. The program, user's guide and example file can be downloaded free of charge and from the World Wide Web "ENDOG dowload area" at http://www.ucm.es/info/prodanim/html/JP_Web.htm#_Endog_3.0:_A. If ENDOG is going to be installed under Windows Seven, user is expected to answer "No" to any query of the setup menu on replacing any .dll file. Otherwise, user can simply

Copy&Paste the ENDOG exceutable file in any folder of the PC running under Windows Seven and directly work with it. ENDOG has been tested on several data sets and results were checked for consistency with alternative software when possible. The authors would appreciate being informed of any detected bug. Although ENDOG has been designed primarily for work with endangered populations and a small sample file is provided with the program, ENDOG can handle very large data files (González-Recio et al., 2007). However, computations on large datasets can take a long time, being limited primarily by the computer processing capability.

1.2 Notice and Disclaimer

Please send bibliographic information, preferably a reprint, about every paper in which ENDOG is used to any author of this document. Please report any errors found to author's address as indicated (preferably to Juan Pablo Gutiérrez by e-mail). We would very much appreciate users submitting their suggestions for improvements to this manual, directly by e-mail, just sending an improved version of the ENDOG User's Guide.

This program is provided 'as-is'. No authors could be held responsible in case of trouble. Although this program has been tested, the authors make no warranty as to the accuracy and functioning of the program. You may distribute this program freely in any format, so long as the following conditions are met: the program remains intact without modification, the help file is included without modification, no fee of any kind is charged.

If the user is interested in being informed of further developments of the program ENDOG please contact the authors by email.

1.3 News and Further Development

Since the paper describing the main characteristics of ENDOG v 3.0 was published in the Journal of Animal Breeding and Genetics (Gutiérrez and Goyache, 2005) we have received a significant number of comments and suggestions from the users. Most of them have lead to slight modifications in the program that have improved its performance leading to renaming the program as ENDOG v 3.2. However, at that time, the included modifications did not affect the user's options except for the writing to disk of the Table f_{ij} in the Fstats submenu. From version 4.2, ENDOG included the possibilities of computing N_e from family size variance and further assessment of the level of completeness of the analysed pedigrees. These improvements together with a less costly handling of the relationship matrix are the basis of ENDOG v4.3. ENDOG 4.6 included the computation of Lacy et al.'s (1996) partial inbreeding coefficients and Gutiérrez et al.'s (2008) individual increase in inbreeding coefficients. The computation of N_e from variances in family sizes and regression coefficients was also improved with respect to previous versions of the program, allowing more generalised approaches. The computation of the Alderson's (1992) Genetic Conservation Index was also included in ENDOG v4.6. ENDOG v4.6 included the possibility of computing most parameters for different reference populations and storing the obtained results in the corresponding Access tables without starting a new session of the program. Version 4.8 includes the modification on the computation of the Cervantes et al.'s (2008) 'realized effective size' proposed by Gutiérrez et al. (2009). This modification computes individual increase in inbreeding on *t-1* generations due to inbreeding is not possible in the first generation in populations with two sexes. Also, version 4.8 allows user to compute N_e from paired increase in conacestry (Cervantes et al., 2010), the assesment of the contributions of predefined subpopulations to total diversity following Caballero and Toro (2002) and a methodology for adjustment of within- and betwen-subpopulations coancestries by subpopulations size (Bartolomé et al., 2010).

Many users have suggested the inclusion in ENDOG of additional parameters. Users are kindly requested to send the authors their own routines (in any programming language) with a (brief) explanation on the interest of including them in future versions of ENDOG. These routines will be appropriately acknowledged in further modifications of this User's Guide.

The compatibility between ENDOG and Widows Vista is poor depending on the particular configuration of this environment in each PC and may give problems to some users. If problems arise, the users are advised to copy the executable file in a folder directly without any other isntalling procedure. Some users are interested in obtain a version of ENDOG running under LINUX. This software has not been programmed to run efficiently in the LINUX desktop environment. Although users have suggetsed to run ENDOG on top of Wine, an open-source compatibility layer for running Windows programs on Unix-based Operating Systems, this has not been tested by the authors. Users are kindly requested to suggest the most appropriate ways to deal with these tasks.

1.4 How to cite ENDOG (v 4.8)

If you wish to cite the use of ENDOG in your publications, we suggest the following citation:

Juan Pablo Gutiérrez and Félix Goyache (2005) A note on ENDOG: a computer program for analysing pedigree information. Journal of Animal Breeding and Genetics, 122: 172-176.

When 'realised' N_e from individual increase in inbreeding (Gutiérrez et al., 2008, 2009) or N_e from paired increase in coancestry (Cervantes et al., 2010) are computed, users are kindly requested to cite the papers in which these procedures are described.

2.-What ENDOG (v4.8) Does

Primary functions carried out by ENDOG are the computation of the individual inbreeding (*F*) (Wright, 1931) and the average relatedness (*AR*) (Gutiérrez et al., 2003; Goyache et al., 2003) coefficients. Information on the completeness of pedigree is also provided. Additionally, ENDOG enables users to compute useful parameters in population genetics such as that described by Boichard et al. (1997) for the number of ancestors explaining genetic variability or those proposed by Robertson (1953) and Vassallo et al. (1996) for the genetic importance of the herds. Moreover, ENDOG can compute F-statistics (Wright, 1978) from genealogical information following Caballero and Toro (2000; 2002). Different approaches to compute effective population size (N_e) from the increase of inbreeding were implemented in the former versions of the program. From ENDOG v4.6 N_e could be computed from family size variances using the formula by Hill (1979). Different approaches to ascertain the genetic contributions of founders or ancestors to a reference population are available including the computation of partial inbreeding coefficients (Lacy et al., 1996) and the Genetic Conservation Index (Alderson, 1992). The present version of ENDOG (4.8) allows computing N_e from individual increase in inbreeding (Gutiérrez et al., 2008) using the modification propossed by Gutiérrez et al. (2009) but also from paired increase in coancestry (Cervantes et al., 2009). Coancestries may be used now to ascertain the contributions of predefined populations to total diversity (Caballero and Toro, 2002) but also they can be adjusted for subpopulation size (Bartolomé et al., 2010) to obtain sound differentiation (F_{ST} and others) or contribution to diversity values.

2.1 Inbreeding and Effective Size

F is defined as the probability that an individual has two identical alleles by descent, and is computed following Meuwissen and Luo (1992). The increase in inbreeding (ΔF) is calculated for each generation by means of the classical formula $\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$, where F_t and F_{t-1} are the average inbreeding at the i_{th} generation. Using ΔF, ENDOG computes the effective population size (N_e) as $N_e = \frac{1}{2\Delta F}$ for each generation having $F_t > F_{t-1}$ to roughly characterise the effect of remote and close inbreeding. N_e is defined as the number of breeding animals that would lead to the actual increase in inbreeding if they contributed equally to the next generation. In small populations with shallow pedigrees, whatever method is used to compute N_e, this parameter fits poorly with real populations giving an overestimate of the actual effective population size (Goyache et al., 2003). To better characterize this, ENDOG gives three additional values of N_e by computing the regression coefficient (*b*) of the individual inbreeding coefficient over: i) the number of full generations traced; ii) the maximum number of generations traced; and iii) the equivalent complete generations, and considering the corresponding regression coefficient as the increase in inbreeding between two generations ($F_t - F_{t-1} = b$), and consequently $N_e = \frac{1}{2b}$. When available information is scarce, these estimations can be useful to inform on the lower using (i), upper (ii) and 'real' (using iii) limits of N_e in the analysed population.

ENDOG also allows computing ΔF for a given reference subpopulation using different regression based approaches. First ENDOG computes the increase in inbreeding as $\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}} \approx \frac{b}{1 - (F_t - b)}$ being F_t the average *F* of the reference subpopulation and b the regression coefficient of the individual inbreeding coefficients over the equivalent complete generations. Additionally the effective size is estimated, following Gutiérrez et al. (2003), from the regression coefficient (*b*) of the inbreeding coefficients over the year of birth in a reference population and computing the increase in inbreeding between two generations as F_t - F_{t-1} = *l* x *b*, where *l* is the average generation interval and F_t is the mean inbreeding in the reference subpopulation. Finally, ENDOG allows computing the approach by Pérez-Enciso (1995) to estimate N_e via a log regression of (1 - F_t) on generation number (where F_t is obtained from $F_t = 1 - (1 - \Delta F)^t$ as $1 - F_t = \left(1 - \frac{1}{2N_e}\right)^t$). When datasets with no discrete generations are analysed, N_e was estimated by a log regression of (1 - F_t) on the date of birth and then divided by the generation interval (Pérez-Enciso, 1995).

2.2 Individual increase in inbreeding and 'realized' N_e

From its version 4.5, ENDOG includes the estimation of effective population sizes from individual increase in inbreeding (ΔF_i) following the approach recently proposed by Gutiérrez et al. (2008) in the form proposed by Gutiérrez et al. (2009). The ΔF_i coefficients are simply computed as $\Delta F_i = 1 - \sqrt[t-1]{1-F_i}$, where F_i is the individual coefficient of inbreeding and t is the 'equivalent complete generations' (Maignel et al., 1996; see section 2.5). This estimate of effective population size ($\overline{N_e}$), called 'realized effective size' by Cervantes et al. (2008), can be computed from $\overline{\Delta F}$, that can be easily computed by averaging the ΔF_is of the n individuals included in a given reference subpopulation, as $\overline{N_e} = \frac{1}{2\overline{\Delta F}}$. Note that this way of computing effective population size is not dependent on the whole reference population mating policy but on the matings carried out throughout the pedigree of each individual (Gutiérrez et al., 2008). Moreover, the modification proposed by Gutiérrez et al. (2009) correct for the fact that self-fertilization is not possible in farm animal populations. Interestingly, a standard error of $\overline{N_e}$ can be computed from the standard deviation of $\overline{\Delta F}$ and the square root of the size (n) of the reference subpopulation as $\sigma_{\overline{\mathrm{N}}\mathrm{e}} = 2\overline{\mathrm{N}}_{\mathrm{e}}^2\sigma_{\Delta\mathrm{F}}\frac{1}{\sqrt{\overline{\mathrm{N}}_{\mathrm{e}}}}$ (see Gutiérrez et al., 2008, for a complete description of the method).

2.3 Paired Increase in Coancestry

Following similar reasonings than in the case of individual increase in inbreeding (ΔF_i), Cervantes et al. (2010) introduced a simple method to estimate effective populations size from increase in coancestry for all pairs of individuals j and k (Δc_{jk}) in a reference subpopulation. This parameter is computed as $\Delta c_{jk} = 1 - \sqrt[\left(\frac{g_j+g_k}{2}\right)]{1-c_{jk}}$, where c_{jk} is the inbreeding value corresponding to an offspring from j and k, and g_j and g_k are the discrete equivalent generation of individuals j and k. Averaging the increase in coancestry for all pairs of individuals in a reference subpopulation, we can estimate a realised effective population size based on coancestries as $\overline{N}_{ec} = \frac{1}{2\overline{\Delta c}}$, that would provide information on the effective size of a population under random mating. A standard error of the $\overline{N}_{ec}$ can be further computed from the standard deviation of these increases in coancestry $\left(\sigma_{\Delta c}\right)$ and the square root of the effective size of the effective number of paired coancestries in the reference sub population as

$$\sigma_{\overline{N}_{ec}} = 2\overline{N}_{ec}^{\,2}\sigma_{\Delta c}\frac{1}{\sqrt{\frac{\overline{N}_{ec}\left(\overline{N}_{ec}-1\right)}{2}}}.$$

2.4 Partial Inbreeding Coefficients

ENDOG includes the possibility of computing partial inbreeding coefficients from a given number of founders or ancestors following Lacy et al. (1996). You can find an example of the usefulness of partial inbreeding coefficients in animal breeding in Man et al. (2007). The partial inbreeding coefficient is the probability that an individual is homozygous (identical by descent) for an allele descended from the specified founder. The sum, across all founders, of the partial inbreeding coefficients for a descendant is equal to the inbreeding coefficient for that individual. A detailed explanation of the method for computing partial inbreeding coefficients can be found on Robert Lacy's website at http://www.vortex9.org/partinbrdef.html. Note that computations on datasets with very dense relationships can take a long time.

2.5 Average Relatedness

The average relatedness coefficient (*AR*) of each individual is defined as the probability that an allele randomly chosen from the whole population in the pedigree belongs to a given animal. *AR* can then be interpreted as the representation of the animal in the whole pedigree regardless of the knowledge of its own pedigree. Computation of *AR* has not been formalized before. It uses an algorithm to obtain a vector **c'** defined as:

$$\mathbf{c'} = (1/n)\,\mathbf{1'A} \qquad \textbf{[1]}$$

A being the numerator relationship matrix of size n x n. On the other hand, the numerator relationship matrix can be obtained from the **P** matrix, where p_{ij} is equal to 1 if j is parent of i, and 0 otherwise, which sets the parents of the animals (Quaas, 1976), by:

$$\mathbf{A} = (\,\mathbf{I} - \tfrac{1}{2}\,\mathbf{P}\,)^{-1}\,\mathbf{D}\,(\,\mathbf{I} - \tfrac{1}{2}\,\mathbf{P'}\,)^{-1} \qquad \textbf{[2]}$$

where **D** is a diagonal matrix with non-zero elements obtained by:

$$d_{ii} = 1 - \tfrac{1}{4}\,a_{jj} - \tfrac{1}{4}\,a_{kk} \qquad \textbf{[3]}$$

j and k being the parents of the individual i.

From **[2]**, $\mathbf{A}\,(\,\mathbf{I} - \tfrac{1}{2}\,\mathbf{P'}\,) = (\,\mathbf{I} - \tfrac{1}{2}\,\mathbf{P}\,)^{-1}\,\mathbf{D}$

Premultiplying by (1/n) **1'**:

$$(1/n)\,\mathbf{1'}\,\mathbf{A}\,(\,\mathbf{I} - \tfrac{1}{2}\,\mathbf{P'}\,) = (1/n)\,\mathbf{1'}\,(\,\mathbf{I} - \tfrac{1}{2}\,\mathbf{P}\,)^{-1}\,\mathbf{D}$$

and using **[1]**:

$$\mathbf{c'}\,(\,\mathbf{I} - \tfrac{1}{2}\,\mathbf{P'}\,) = (1/n)\,\mathbf{1'}\,(\,\mathbf{I} - \tfrac{1}{2}\,\mathbf{P}\,)^{-1}\,\mathbf{D}$$

Multiplying **c'** into the parenthesis and isolating **c'**:

$$\mathbf{c'} = (1/n)\,\mathbf{1'}\,(\,\mathbf{I} - \tfrac{1}{2}\,\mathbf{P}\,)^{-1}\,\mathbf{D} + \tfrac{1}{2}\,\mathbf{c'}\,\mathbf{P'} \qquad \textbf{[4]}$$

The advantages of using *AR* are: a) the computational cost to calculate *AR* coefficients is similar to that for the computation of the numerator relationship matrix, since both use common algorithms; b) *AR* of a founder indicates its genetic contribution to the population; c) *AR* coefficients can also be used as a measure of inbreeding of the population, as it takes into account both inbreeding and coancestry coefficients; d) *AR* can be used as an index to maintain the initial genetic stock by using as breeding animals those with the lowest AR value; and e) *AR,* as an alternative or complement to

F, can be used to predict the long-term inbreeding of a population because it takes into account the percentage of the complete pedigree originating from a founder at population level. In addition, *AR* can be used to compute the effective size of the founder population as the inverse of the sum of the square AR coefficients across founder animals.

2.6 Pedigree Completeness

At the same time as the computation of F and AR coefficients, ENDOG computes for each individual the number of fully traced generations, the maximum number of generations traced and the equivalent complete generations for each animal in the pedigree data. The first is defined as those separating the offspring of the furthest generation where the 2^g ancestors of the individual are known. Ancestors with no known parent are considered as founders (generation 0). The second is the number of generations separating the individual from its furthest ancestor. The equivalent complete generations is computed as the sum over all known ancestors of the terms computed as the sum of $(1/2)^n$ where n is the number of generations separating the individual to each known ancestor (Maignel et al., 1996).

ENDOG includes explicit information on the quality of pedigree, namely the description of the completeness of each ancestor in the pedigree to the 5th parental generation and MacCluer et al.'s (1983) index of completeness.

2.7 Selection of Ancestors

Using ENDOG we can assess the concentration of the origin of both animals and genes by calculating the following parameters: a) effective number of founders (f_e); b) effective number of ancestors (f_a) (Boichard et al., 1997); and c) effective number of founder herds (f_h). From the version 4.5 of ENDOG the computation of the parameter 'founder genome equivalents' (f_g) (Ballou and Lacy, 1995) is explicitly included in the program. The first is defined as the number of equally contributing founders that would be expected to produce the same genetic diversity as in the population under study. This is computed as: $f_e = 1 \Big/ \sum_{k=1}^{f} q_k^2$ where q_k is the probability of gene origin of the k ancestor. As explained above, this is computed using the AR coefficients of founder individuals; parameter f_e, as computed by ENDOG would be equivalent to that computed following James (1972) or Lacy (1989) if the reference population used was the whole pedigree. Parameter f_a is the minimum number of ancestors, not necessarily founders, explaining the complete genetic diversity of a population. This parameter complements the information offered by the effective number of founders accounting for the losses of genetic variability produced by the unbalanced use of reproductive individuals producing bottlenecks. This parameter is computed in a similar way to the effective number of founders: $f_a = 1 \Big/ \sum_{j=1}^{a} q_j^2$ where q_j is the marginal contribution of an ancestor *j*, which is the genetic contribution made by an ancestor that is not explained by other ancestors chosen before. The last two parameters are initially computed by ENDOG taking as reference population those animals in the pedigree with both parents known. However they can be recomputed after choosing a particular reference population. The effective number of herds is simply computed as the inverse of the summed squared of the sum of the contributions of Boichard et al.'s (1997) ancestors into each herd. Parameter f_g (founder genome equivalents; Ballou and Lacy, 1995) can be defined as the number of founders that would be expected to produce the same genetic diversity as in the population under study if the founders were equally represented and no lost of alleles occurred. Following Caballero and Toro (2000), parameter f_g was obtained by the inverse of twice the average coancestry of the individuals included in a pre-defined reference population.

2.8 Genetic Conservation Index

ENDOG includes the computation of the genetic conservation index (*GCI*; Alderson, 1992) for each of the individuals of the analysed population. The index is computed from the genetic contributions of all the identified founders as $GCI = \frac{1}{\sum p_i^2}$ where p_i is the proportion of genes of founder *i* in the pedigree of an animal. The Alderson's (1992) index is based on the assumption that the objective of a conservation program is to retain the full range of alleles possessed by the base population. In this respect, the ideal individual would receive equal contributions from all the founder ancestors in the population and, consequently, the higher the *GCI* value the higher the values of an animal for conservation (Alderson, 1992). The index has limitations such as not accounting for any concentration of breeding to non-founder animals in subsequent generations in a pedigree (Alderson, 1992).

2.9 F-statistics

ENDOG can be used to infer structure of population(s) from pedigree information. ENDOG can compute Nei's minimum distance (Nei, 1987) and F statistics (Wright, 1978) for each subpopulation we can define (i.e. sex, areas, herds, etc.). Wright's F statistics are computed following Caballero and Toro (2000, 2002). These authors have formalized the pedigree tools necessary for the analysis of genetic differentiation in subdivided populations starting with the average pairwise coancestry coefficient ($f_{i\,j}$) between individuals of two subpopulations, *i* and *j,* of a given metapopulation including all $N_i \times N_j$ pairs. For a given subpopulation *i*, the average coancestry, the average self-coancestry of the N_i individuals and the average coefficient of inbreeding would be, $f_{ii,}$, s_i, $F_i = 2s_i - 1$ respectively. The average distance between individuals of subpopulations *i* and *j* would be $D_{ij} = \left[\left(s_i + s_j\right)/2\right] - f_{ij}$. Note that this distance has been called as *kinship distance* (D_k) for molecular coancestry, by Eding and Meuwissen (2001) From these parameters and the corresponding means for the entire metapopulation Caballero and Toro (2000, 2002) obtained the genetic distance between subpopulations *i* and *j* (Nei's minimum distance; Nei 1987) as $\mathbf{D_{ij}} = D_{ij} - \left[\left(D_{ii} + D_{jj}\right)/2\right] = \left[\left(f_{ii} + f_{jj}\right)/2\right] - f_{ij}$, and its average over the entire metapopulation as $\overline{D} = \frac{\sum_{i,j=1}^{n} D_{ij} N_i N_j}{N_T^2}$, that are the equations (3) and (4) of Caballero and Toro (2002). Finally, the Wright's (1978) *F*- statistics are obtained as $F_{IS} = \frac{\tilde{F} - \bar{f}}{1 - \bar{f}}$,

$F_{ST} = \dfrac{\bar{f} - \tilde{f}}{1 - \tilde{f}} = \dfrac{\bar{D}}{1 - \tilde{f}}$ and $F_{IT} = \dfrac{\tilde{F} - \tilde{f}}{1 - \tilde{f}}$, where $\tilde{f}$ and $\tilde{F}$ are, respectively, the mean coancestry and the inbreeding coefficient for the entire metapopulation, and, $\bar{f}$ the average coancestry for the subpopulation, so that $(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$.

2.10 Contribution of Subpopulations to Total Diversity

ENDOG v4.8 allows assessing genetic contributions of subpopulations to total diversity following Caballero and Toro (2002). The average coancestry (Malècot, 1948), $\bar{f}$, over an entire metapopulation of N_T individuals consisting of n subpopulations, subpopulation i with N_i breeding individuals, is:

$$\bar{f} = \frac{\sum_{i,j=1}^{n} f_{ij} N_i N_j}{N_T^2} = \frac{\sum_{i=1}^{n} f_{ii} N_i}{N_T} - \bar{\mathbf{D}} = \sum_{i=1}^{n} \frac{N_i}{N_T} \left[f_{ii} - \frac{\sum_{j=1}^{n} \mathbf{D}_{ij} N_j}{N_T} \right]$$

, being f_{ij} the average pairwise coancestry between individuals of subpopulations i and j, including all $N_i \times N_j$ pairs and f_{ii} the average pairwise coancestry within subpopulation i and where $\boldsymbol{D}_{ij}$ is the Nei's minimum genetic distance (Nei ,1987) between subpopulations i and j computed as $\boldsymbol{D}_{ij} = [(f_{ii} + f_{jj})/2] - f_{ij}$. From the formula above it can be noted that $\bar{f}$ is dependent on the within-subpopulation coancestry (first term in the brackets) and the average distance among subpopulations (second term in the braquets). Proportional contribution of each subpopulation to the global coancestry can be computed as the average coancestry of the subpopulation minus its average distance with all the others.

2.11 Adjustment for Subpopulation Size

The within-subpopulation coancestry is affected by the subpopulation size. Self-coancestries $\left[s_i = \tfrac{1}{2}(1 + F_i)\right]$ will have more importance in small populations. In an extreme case of a group of non-inbred and non-related individuals, off-diagonal will be null and only N_i self-coancestries with value ½ will take part in the mean. Therefore, in this case average within-subpopulation coancestry will be $f_{ii} = \dfrac{N_i \tfrac{1}{2}}{N_i^2} = \dfrac{1}{2N_i}$. Obviously, as the subpopulation size decreases, the within-subpopulation coancestry tends to increase. ENDOG implements the method described in Bartolomé et al. (2010) to adjust the within-subpopulation coancestry can be adjusted via extrapolation to a given subpopulation size M as:

$$f_{ii}^{M} = f_{ii}^{N_i} - \frac{s_i}{N_i} + \frac{s_i}{M} = f_{ii}^{N_i} + \frac{(N_i - M) s_i}{M N_i}$$

, being s_i the mean self-coancestry in the subpopulation, and $f_{ii}^{N_i}$ and f_{ii}^{M} the respective within-subpopulation coancestry means for the original sample size and the desired sample size of M.

2.12 Genetic Importance of the Herds

At herd level, besides the effective number of herds, ENDOG computes the genetic importance of the herds in a population as the contribution of the herds with reproductive males to the population (Vassallo et al., 1986). Using this methodology the herds are classified as: i) nucleus herds, if breeders use only their own males, never purchase males but sell them; ii) multiplier herds, when breeders use purchased males and also sell males; and iii) commercial herds if they use purchased males and never sell males. Additionally, ENDOG computes the inverse of the probability that two animals taken at random in the population have their parent in the same herd for each path to know the effective number of herds supplying fathers (H_S), grandfathers (H_{SS}) and great-grandfathers (H_{SSS}) (Robertson, 1953).

2.13 Generation Intervals

At population or subpopulation level, ENDOG computes both the generation intervals, defined as the average age of parents at the birth of their progeny kept for reproduction (James, 1977), and the average age of parents at the birth of their offspring (used for reproduction or not). Both parameters are computed for the 4 pathways (father- son, father - daughter, mother - son and mother - daughter).

2.14 N_e estimation based on family size variance

Finally, ENDOG computes effective size (*Ne*) from the variances of family sizes as

$$\frac{1}{N_e}=\frac{1}{16ML}\left[2+\sigma^2_{mm}+2\left(\frac{M}{F}\right)\mathrm{cov}(mm,mf)+\left(\frac{M}{F}\right)^2\sigma^2_{mf}\right]+\frac{1}{16FL}\left[2+\left(\frac{F}{M}\right)^2\sigma^2_{fm}+2\left(\frac{F}{M}\right)\mathrm{cov}(fm,ff)+\sigma^2_{ff}\right]$$

(Hill, 1979) where M and F are the number of male and female individuals born or sampled for breeding each time period, L the average generation interval σ^2_{mm} and σ^2_{mf} are the variances of the male and female offspring of a male, σ^2_{fm} and σ^2_{ff} are the variances of the male and female offspring of a female, and $\mathrm{cov}(mm,mf)$ and $\mathrm{cov}(fm,ff)$ the respective covariances. Note also that the family size of a parent (male or female) consists in its number of sons and daughters kept for reproduction (James, 1977). As explained below (see section 3.2.19), ENDOG can compute N_e on a predefined reference population but also on the cohorts formed by the individuals born in a period time of approximately equal to the generation interval.

3.- How to Use ENDOG (v 4.8)

You can install ENDOG following a setup menu. **NOTE** that the setup menu is in Spanish and "**Salir**" means '**Exit**'. Please do not click on the "**Salir**" box if you want to install ENDOG!!!!

3.1 Input Files

ENDOG has been designed to avoid much preparation of input files. ENDOG accepts xls files (from Microsoft Excel worksheets) or dbf files. Files with dbf format must be used in datasets larger than the limit of rows of Excel (65,536). Columns (or fields) are not supposed to be in a given order and no strict identification of the columns is needed.

WARNING!! Boxes in the Excel worksheet must be written without any comment or special format to ensure a correct operation of ENDOG. Moreover, the format of the identification of the individuals must be consistent with that of the identification of the fathers and the identification of the mothers. The format of the column (or field) including birth dates must not be formatted as character. Be particularly careful to avoid unaccounted additional rows or columns in the Excel worksheet used as input.

Excel does not work with birthdates before 1900. In consequence, no Excel spreadsheets can be used in these conditions and .dbf files must be used as input data.

At the beginning of a session ENDOG will ask for file containing the data set and, if xls, for the specific worksheet in which data resides. After that, the program will ask if records are renumbered and ordered sequentially (from 1 to n, the older the lower number) and, later, for the selection of the column (or field) containing the identification of the individuals, the identification of the fathers, the identification of the mothers, and the sex of the individuals. Numbering and ordering individuals is recommended but, in fact, individuals can be identified in any way (using numbers, characters or both). However, identifications used for individuals must be consistent with those used for parents.

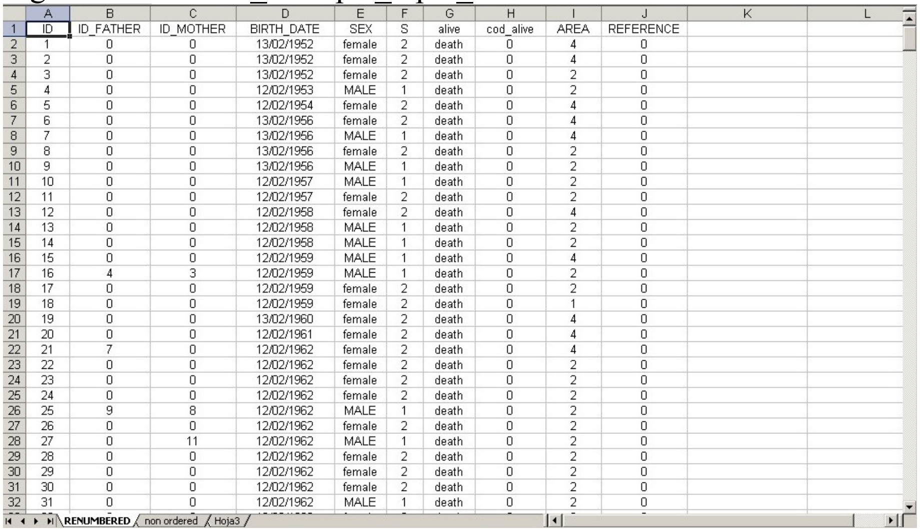
If records are not renumbered and sequentially ordered, ENDOG will ask for the column (or field) in which the individuals' birth date is included to proceed to order data. Note that if records are not ordered all the individuals must have a consistent birth date. Where actual birth dates are not available users must create virtual (but consistent) birth dates. Since computation of generation intervals can be done on columns (or fields) including unknown birth dates we recommend that the inclusion of approximate birth dates is done in a different column (or field) to that including actual birth dates. Later the true birth date information is included e.g. in computing the generation interval (while the use of approximate birth date in that context would results in wrong estimates). As explained in the next section, it is not an absolute requirement, but it is highly recommended that records were renumbered and sequentially ordered to avoid errors in computations. Identification of the parents must be consistent with that of the individuals. Sex must be coded as 1 for males and 2 for females. Despite these shortcomings, the input file can have any other columns (or fields) in any format (character, date, numerical or other). These columns can contain any other data to identify the individuals, the identification of the herd or population corresponding to the individual or any other. The inclusion of a column with the birth date of the animals in the input file is highly recommended because this information will be needed for some procedures.

Users interested in computing parameter using a particular reference population must include in the input file a column (or field) in which the animals forming the reference population were identified using the number 1.

An xls file called 'ENDOG_example_input_file.xls' is provided with the program. You can find two different worksheets called 'renumbered' and 'non ordered'. These are basically the same but the former has the individuals numerically identified and sequentially ordered and the latter has not. When data is renumbered, unknown parents must be identified with 0; otherwise unknown parents can be identified as blank space or 0.

Identifications of the columns are ID, ID_FATHER, ID_MOTHER, BIRTH_DATE, SEX, S, alive, cod_alive, AREA and REFERENCE. Names of the columns are self-informative on the content. To show that input file can have columns (or fields) with the same information in different formats, SEX and alive are shown recoded numerically. Note that ENDOG will only work with the sex identification if it is numerically coded. The AREA column can be used for subpopulation or herd in the corresponding procedures. In any case, herds or subpopulations can be identified numerically. REFERENCE is the column containing the information on the reference population to be used for computing some parameters instead of the default population used for ENDOG (all the individuals with both parents known).

Figure 1: ‘ENDOG_example_input_file.xls’



	A	B	C	D	E	F	G	H	I	J
1	ID	ID_FATHER	ID_MOTHER	BIRTH_DATE	SEX	S	alive	cod_alive	AREA	REFERENCE
2	1	0	0	13/02/1952	female	2	death	0	4	0
3	2	0	0	13/02/1952	female	2	death	0	4	0
4	3	0	0	13/02/1952	female	2	death	0	2	0
5	4	0	0	12/02/1953	MALE	1	death	0	2	0
6	5	0	0	12/02/1954	female	2	death	0	4	0
7	6	0	0	13/02/1956	female	2	death	0	4	0
8	7	0	0	13/02/1956	MALE	1	death	0	4	0
9	8	0	0	13/02/1956	female	2	death	0	2	0
10	9	0	0	13/02/1956	MALE	1	death	0	2	0
11	10	0	0	12/02/1957	MALE	1	death	0	2	0
12	11	0	0	12/02/1957	female	2	death	0	2	0
13	12	0	0	12/02/1958	female	2	death	0	4	0
14	13	0	0	12/02/1958	MALE	1	death	0	2	0
15	14	0	0	12/02/1958	MALE	1	death	0	2	0
16	15	0	0	12/02/1959	MALE	1	death	0	4	0
17	16	4	3	12/02/1959	MALE	1	death	0	2	0
18	17	0	0	12/02/1959	female	2	death	0	2	0
19	18	0	0	12/02/1959	female	2	death	0	1	0
20	19	0	0	13/02/1960	female	2	death	0	4	0
21	20	0	0	12/02/1961	female	2	death	0	4	0
22	21	7	0	12/02/1962	female	2	death	0	4	0
23	22	0	0	12/02/1962	female	2	death	0	2	0
24	23	0	0	12/02/1962	female	2	death	0	2	0
25	24	0	0	12/02/1962	female	2	death	0	2	0
26	25	9	8	12/02/1962	MALE	1	death	0	2	0
27	26	0	0	12/02/1962	female	2	death	0	2	0
28	27	0	11	12/02/1962	MALE	1	death	0	2	0
29	28	0	0	12/02/1962	female	2	death	0	2	0
30	29	0	0	12/02/1962	female	2	death	0	2	0
31	30	0	0	12/02/1962	female	2	death	0	2	0
32	31	0	0	12/02/1962	MALE	1	death	0	2	0

3.2 Output Files

Most results of ENDOG are written in a Microsoft ACCESS file named Gener.mdb to facilitate further use. Results of each analysis are written to the corresponding Table within Gener.mdb file. However, ENDOG also presents summary results for on screen viewing after most analyses. These summary results are written in their corresponding txt files with delimited format, to allow their editing using most common spreadsheet programmes. The names of the ACCESS tables and txt files containing the results of the computations are usually self informative on the content and are listed in Table 1.

Table 1: List of the result files obtained using ENDOG

Procedure	ACCESS table	txt results file	Description
Initial check		error.txt	List of errors found in the pedigree
Default computations	Midef		Gives F, AR, and number of generations for each individual in the dataset.
		Coan.txt Coan_Ref.txt	They give the coancestry matrix for the whole population or the coancestry matrix between the individuals of a predefined references subpopulation and are written to disk after clicking on the corresponding box.
	Inbreed_?		Gives F for the number of generations (_?) fitted by user.
	GCI		Gives the Alderson’s Genetic Conservation Index
		HighInbred.txt	This file gives the frequency of matings between close relatives.
Population Menu			
Inbreeding per	PorG	Populat.txt	Mean values of F, AR and N_e for

Generations submenu	PorC		the whole population and each generation traced
		Ne_IncInb.txt	Effective Population Size computed via individual increase in inbreeding, via paired increase in coancestry (if requested) and via regression approaches for a given reference subpopulation.
Pedigree Content Submenu	PediCont PediContRef		Give, respectively, the frequency of the contribution of given ancestor in the pedigree to the 5th parental generation for the whole pedigree or for a given reference population.
	PedKnow PedKnowRef		Give the MacCluer et al.'s (1983) completeness index for the whole pedigree or for a given reference population.
Founders submenu	Founders FoudersRef Ancestors AncestorsRef RebaFund RebaFundRef	Founders.txt Ancestor.txt	Contribution of founders, Boichard et al.'s (1997) ancestor's and founder herds to the population for the whole population or for a predefined reference population.
Partial Inbreeding submenu	F_? F_a_? F_c_?		Partial inbreeding coefficients for each individual in a reference subpopulation due to a given founder or ancestor.
Generation Intervals submenu	GenInterv GeIntRef		Average generation intervals and reproductive ages for each path parent –son, for the whole population or for a given reference population.
		NeGenInterv.txt NeGenInterv_Ref.txt	They give the N_e obtained from regression on the birth date and N_e obtained from Log-regression on the birth date for the whole population or for a given reference population.
Offspring Analysis submenu	NeOffs_Year NeOffs_Gen OffsNeRef		They give estimates of effective population size (N_e) by temporal periods or for a predefined reference population
Subpopulations submenu	fij AverDist DistNei Fis_Fsts	Fij_?.txt MatFst_?.txt AverDist.txt DistNei.txt Fis_Fsts.txt	Paired average coancestry and Average, Nei and Fst distance values for each defined subpopulation. If the number of subpopulations exceeds the maximum number of columns fitted in the ACCESS tables the distance matrices are saved to disk in txt format.

	GenEquiv	This table gives the mean equivalent to complete generations within- and between-subpopulations.
	GainLoss_?	This table gives the contributions of each subpopulation to within-, between- and total gene diversity.
Individual Menu		
Coancestry submenu	Parent	Coancestry values of a key individual with all the individuals of the other sex in the dataset
Herds Menu		
Herds structure submenu	HerStr StrHer Robert	Give information on the genetic importance of each herd in the population, summary of this information and Robertson (1953) statistics

3.2.1. The error.txt file

When ENDOG detects some inconsistencies in the input data, these are written to a txt file and procedures are stopped.

Figure 2: Two examples of error.txt file

Example 1:

```
The field identifying the animal is not the same type than parents.
```

Example 2:

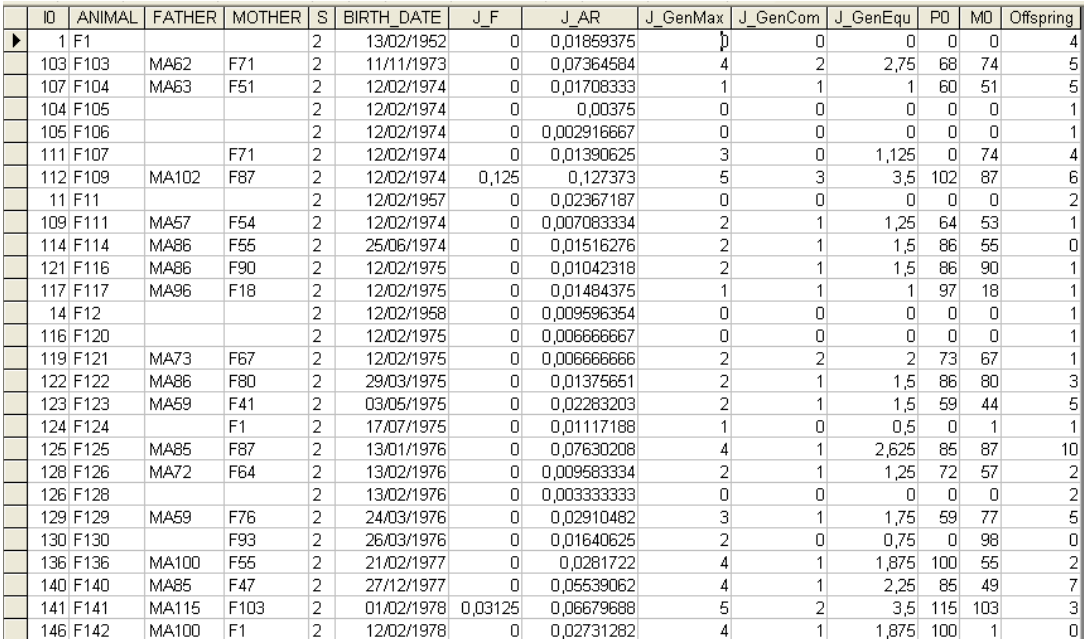
```
The animal  17 has an identification number lower than its father  32
The animal  32 appears as the father of  17 when it is supposed to be
female
The animal  32 has an identification number lower than its mother  0
```

3.2.2. The MiDef Table

This Table is produced by default after the input file is accepted by ENDOG. Besides the identification of the individual, fathers, mothers and birth date (when the data set is not previously renumbered), users will obtain 6 parameters for each individual: J_F (which is the individual inbreeding coefficient), J_AR (which is the individual average relatedness coefficient), J_GenMax (which is the maximum number of generations traced), J_GenCom (which is the number of full generations traced), J_GenEqu (which is the equivalent complete generations), J_AF (individual increase in inbreeding) and offspring (which is offspring size –regardless the sex- of the individual)

If individuals in the file used for input are not renumbered and sequentially ordered MiDef includes three new columns containing consistently renumbered and sequentially ordered identities for both the individuals (I0), the individuals acting as fathers (P0) and the individuals acting as mothers (M0). Most users export these renumbered identities to an Excel Worksheet for subsequent runs of ENDOG. Please note that the column I0 must be renamed (i.e. as ID) to allow re-run of Excel data input.

Figure 3: MiDef table obtained from the Gener.mdb results file



	I0	ANIMAL	FATHER	MOTHER	S	BIRTH_DATE	J_F	J_AR	J_GenMax	J_GenCom	J_GenEqu	P0	M0	Offspring
▶	1	F1			2	13/02/1952	0	0,01859375	0	0	0	0	0	4
	103	F103	MA62	F71	2	11/11/1973	0	0,07364584	4	2	2,75	68	74	5
	107	F104	MA63	F51	2	12/02/1974	0	0,01708333	1	1	1	60	51	5
	104	F105			2	12/02/1974	0	0,00375	0	0	0	0	0	1
	105	F106			2	12/02/1974	0	0,002916667	0	0	0	0	0	1
	111	F107		F71	2	12/02/1974	0	0,01390625	3	0	1,125	0	74	4
	112	F109	MA102	F87	2	12/02/1974	0,125	0,127373	5	3	3,5	102	87	6
	11	F11			2	12/02/1957	0	0,02367187	0	0	0	0	0	2
	109	F111	MA57	F54	2	12/02/1974	0	0,007083334	2	1	1,25	64	53	1
	114	F114	MA86	F55	2	25/06/1974	0	0,01516276	2	1	1,5	86	55	0
	121	F116	MA86	F90	2	12/02/1975	0	0,01042318	2	1	1,5	86	90	1
	117	F117	MA96	F18	2	12/02/1975	0	0,01484375	1	1	1	97	18	1
	14	F12			2	12/02/1958	0	0,009596354	0	0	0	0	0	1
	116	F120			2	12/02/1975	0	0,006666667	0	0	0	0	0	1
	119	F121	MA73	F67	2	12/02/1975	0	0,006666666	2	2	2	73	67	1
	122	F122	MA86	F80	2	29/03/1975	0	0,01375651	2	1	1,5	86	80	3
	123	F123	MA59	F41	2	03/05/1975	0	0,02283203	2	1	1,5	59	44	5
	124	F124		F1	2	17/07/1975	0	0,01117188	1	0	0,5	0	1	1
	125	F125	MA85	F87	2	13/01/1976	0	0,07630208	4	1	2,625	85	87	10
	128	F126	MA72	F64	2	13/02/1976	0	0,009583334	2	1	1,25	72	57	2
	126	F128			2	13/02/1976	0	0,003333333	0	0	0	0	0	2
	129	F129	MA59	F76	2	24/03/1976	0	0,02910482	3	1	1,75	59	77	5
	130	F130		F93	2	26/03/1976	0	0,01640625	2	0	0,75	0	98	0
	136	F136	MA100	F55	2	21/02/1977	0	0,0281722	4	1	1,875	100	55	2
	140	F140	MA85	F47	2	27/12/1977	0	0,05539062	4	1	2,25	85	49	7
	141	F141	MA115	F103	2	01/02/1978	0,03125	0,06679688	5	2	3,5	115	103	3
	146	F142	MA100	F1	2	12/02/1978	0	0,02731282	4	1	1,875	100	1	0

3.2.3. The Coan.txt files

Clicking on the corresponding box on the right-hand side of the main screen of ENDOG allows user to write to disk the whole relationship matrix for the analyzed population (written in the Coan.txt file) or the coancestry matrix for the individuals included in a given reference population (written in the Coan_Ref.txt file) Both the Coan.txt and the Coan_Ref.txt files have three columns separated by spaces (see Figure 4); the two first columns consist in the identification of each pair of individuals and the third the corresponding coancestry value.

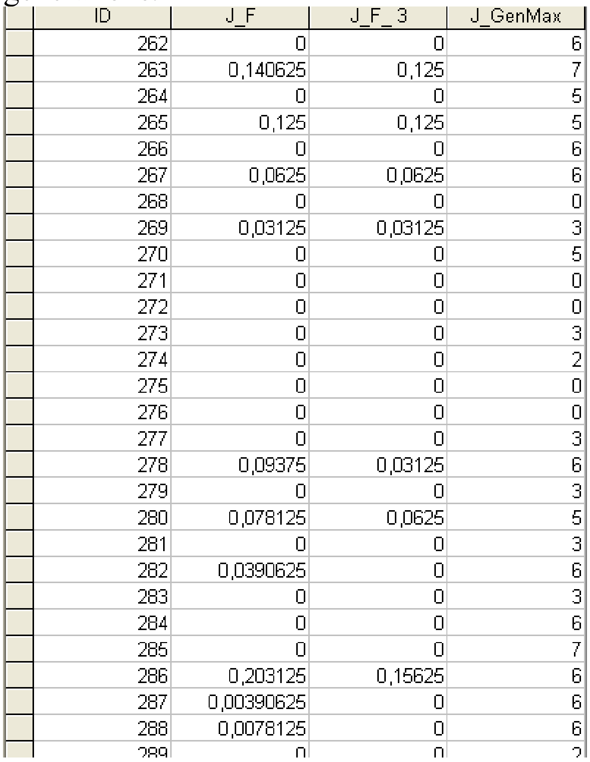
Figure 4: Coan.txt file obtained from the example input file

```
     1     1 0.50000000
     2     1 0.00000000
     2     2 0.50000000
     3     1 0.00000000
     3     2 0.00000000
     3     3 0.50000000
     4     1 0.00000000
     4     2 0.00000000
     4     3 0.00000000
     4     4 0.50000000
     5     1 0.00000000
     5     2 0.00000000
     5     3 0.00000000
     5     4 0.00000000
     5     5 0.50000000
     6     1 0.00000000
     6     2 0.00000000
     6     3 0.00000000
     6     4 0.00000000
     6     5 0.00000000
     6     6 0.50000000
     7     1 0.00000000

 .........................
     600     570 0.00000000
     600     571 0.04589844
     600     572 0.00000000
     600     573 0.21582030
     600     574 0.00097656
     600     575 0.00097656
     600     576 0.10156250
```

```
     600      577 0.04541016
     600      578 0.18261720
     600      579 0.00000000
     600      580 0.00000000
     600      581 0.00000000
     600      582 0.01513672
     600      583 0.00000000
     600      584 0.00000000
     600      585 0.00000000
     600      586 0.00000000
     600      587 0.01757813
     600      588 0.01757813
     600      589 0.00000000
     600      590 0.07910156
     600      591 0.00000000
     600      592 0.00000000
     600      593 0.07910156
     600      594 0.04541016
     600      595 0.00000000
     600      596 0.05761719
     600      597 0.05664063
     600      598 0.20117190
     600      599 0.11071780
     600      600 0.53125000
```

3.2.4. The Inbreed_? Tables

Figure 5: Inbreed_? Table obtained from the example input file after calculating 3 generations.



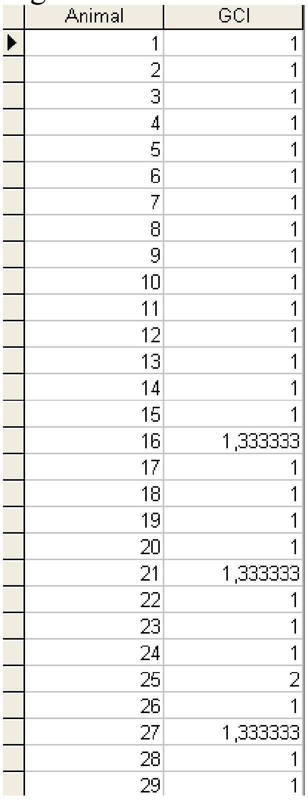
ID	J_F	J_F_3	J_GenMax
262	0	0	6
263	0,140625	0,125	7
264	0	0	5
265	0,125	0,125	5
266	0	0	6
267	0,0625	0,0625	6
268	0	0	0
269	0,03125	0,03125	3
270	0	0	5
271	0	0	0
272	0	0	0
273	0	0	3
274	0	0	2
275	0	0	0
276	0	0	0
277	0	0	3
278	0,09375	0,03125	6
279	0	0	3
280	0,078125	0,0625	5
281	0	0	3
282	0,0390625	0	6
283	0	0	3
284	0	0	6
285	0	0	7
286	0,203125	0,15625	6
287	0,00390625	0	6
288	0,0078125	0	6
289	0	0	2

ENDOG v4.8 allows users to determine both recent and remote inbreeding in a population. Clicking on the box 'Compute recent inbreeding' that the user can find in the main screen of ENDOG, a dialog box will appear asking the user to enter the number of generations to be considered for the computation of F. Results of successive computations will be stored in different Tables with names including the number of generations calculated (in the figure below 3 generations). Each Inbreed_? table will show the actual inbreeding of an individual computed using the whole pedigree (J_F), that computed using only a predefined number of generations (J_F_?) and the maximum number of generations traced for each individual (J_GenMax).

3.2.5. The GCI table

The genetic conservation index (*GCI*) is given in GCI table for each of the individuals in the analysed population (Animal).

Figure 6: GCI Table obtained from the Gener.mdb results file



Animal	GCI
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1,333333
17	1
18	1
19	1
20	1
21	1,333333
22	1
23	1
24	1
25	2
26	1
27	1,333333
28	1
29	1

3.2.6. The Populat.txt file

This file complements the MiDef table described above and contains average values for the main population parameters computed before. It is generated when user clicks on the Inbreeding per Generations submenu.

Figure 7: Populat.txt file obtained from the example input file

```
RESULTS RESUME:

Number of animals:  600
Mean Inbreeding:3,63%
Mean Average Relatedness: 4,02%

Mean Maximum Generations: 3,60
Increase  in  Inbreeding  by  Maximum  Generation:1,41%  =>  Effective
Population Size:   35,51
Mean Complete Generations: 1,32
Increase  in  Inbreeding  by  Complete  Generation:4,06%  =>  Effective
Population Size:   12,31
Mean Equivalent Generations: 2,11
Increase in Inbreeding by Equivalent  Generation:2,86% => Effective
Population Size:   17,50.
```

3.2.7. The HighInbred.txt file

This file informs on the absolute and relative frequency of matings between close relatives that are recorded in the pedigree. It is generated when user clicks on the ‘Highly Inbred matings’ box of the main screen of ENDOG.

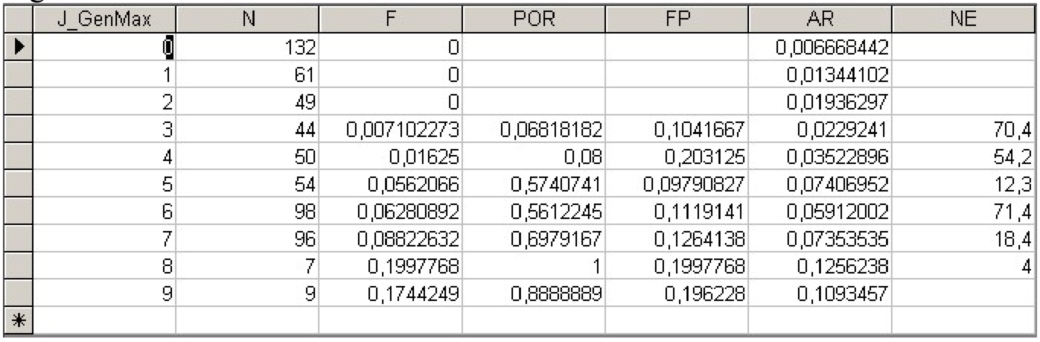
Figure 8: HighInbred.txt file obtained from the example input file

```
 1( 0,17%) matings between full sibs.
 17( 2,83%) matings between half sibs.
 17( 2,83%) matings parent-offspring.
```

3.2.8. The PorG and PorC Tables

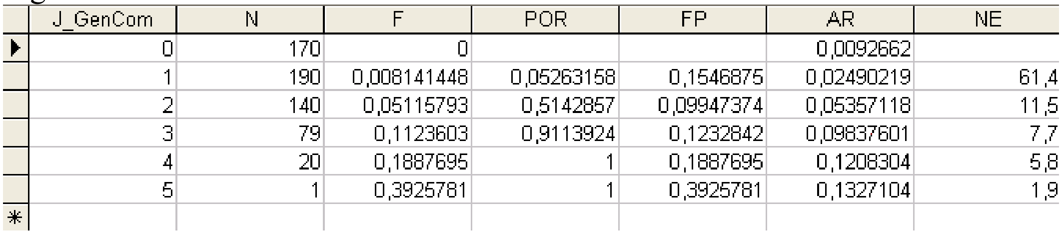
These tables are produced by using the Inbreeding per Generations submenu and are basically the same but showing the information by number of full generations traced (PorC) or by maximum number of generations traced (PorG). Names of fields are: J_GenMax (or J_GenCom) which is the number of generations traced, N (individuals per generation), F (average inbreeding), POR (percent inbred individuals), FP (average inbreeding for inbred individuals, AR (mean average relatedness) and, finally, N_e (effective size) if $F_t > F_{t-1}$.

Figure 9: PorG table obtained from the Gener.mdb results file



J_GenMax	N	F	POR	FP	AR	NE
0	132	0			0,006668442	
1	61	0			0,01344102	
2	49	0			0,01936297	
3	44	0,007102273	0,06818182	0,1041667	0,0229241	70,4
4	50	0,01625	0,08	0,203125	0,03522896	54,2
5	54	0,0562066	0,5740741	0,09790827	0,07406952	12,3
6	98	0,06280892	0,5612245	0,1119141	0,05912002	71,4
7	96	0,08822632	0,6979167	0,1264138	0,07353535	18,4
8	7	0,1997768	1	0,1997768	0,1256238	4
9	9	0,1744249	0,8888889	0,196228	0,1093457	

Figure 10: PorC table obtained from the Gener.mdb results file



J_GenCom	N	F	POR	FP	AR	NE
0	170	0			0,0092662	
1	190	0,008141448	0,05263158	0,1546875	0,02490219	61,4
2	140	0,05115793	0,5142857	0,09947374	0,05357118	11,5
3	79	0,1123603	0,9113924	0,1232842	0,09837601	7,7
4	20	0,1887695	1	0,1887695	0,1208304	5,8
5	1	0,3925781	1	0,3925781	0,1327104	1,9

3.2.9. The Ne_IncInb.txt file

This file complements the MiDef table and the Populat.txt file and contains estimates of N_e computed via individual increase in inbreeding (Gutiérrez et al., 2008) and via regression on equivalent generations for a given subpopulation. It is generated when user clicks on the box asking for the definition of a reference population in the screen showing population statistics on average inbreeding. From ENDOG v4.8 this file contains the effective population size estimated via paired increase in coancestry if requested by the user.

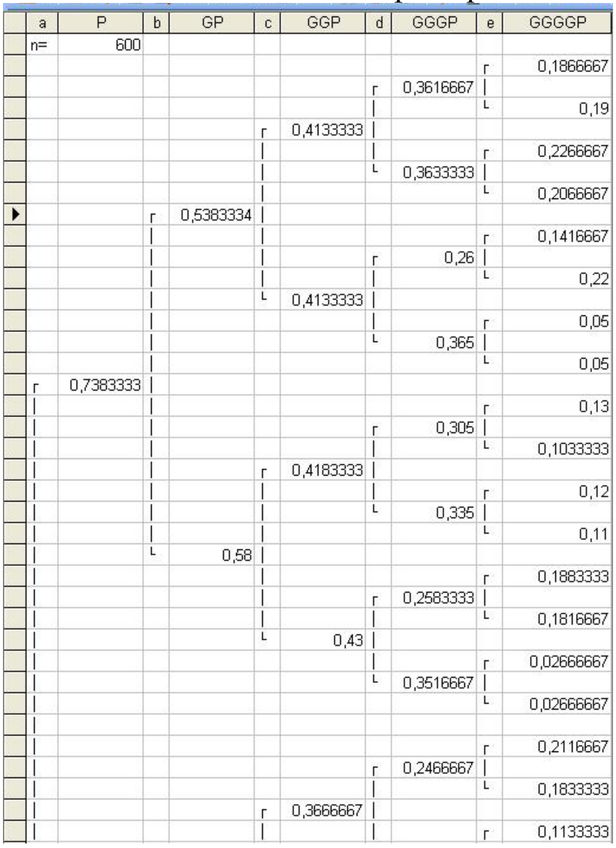
Figure 11: Ne_IncInb.txt file obtained from the example input file

```
Effective Population Size computed via individual increase in
inbreeding =    37,2850
 s.d.=    3,2129
 Nº individuals =  401
 Effective Population Size computed via regression on equivalent
generations =    15,2860
```

3.2.10. The PediCont Tables

The PediCont (Pedicont and PediContRef) Tables are produced by using the Pedigree Content submenu. The PediCont table is generated by default for the whole population whilst the PediContRef table is generated by clicking on the button asking the user for a predefined reference population. They give in separate ‘trees’ for the male and female paths the contribution of each ancestor in the pedigree to the 5th parental generation (first parental generation, fathers; second parental generation, grandfathers; and so on). Each subdivision in branches includes the male ancestry (above in the branch) and the female ancestry (below in the branch). The column P corresponds to parents, GP to grandparents, GGP, great-grandparents, and so on.

Figure 12: PediCont Table obtained from the example input file

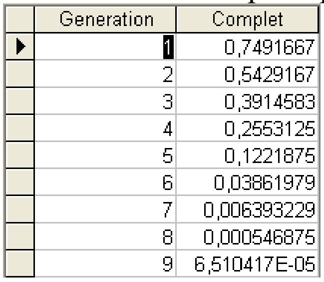


	a	P	b	GP	c	GGP	d	GGGP	e	GGGGP
	n=	600								
									┌	0,1866667
							┌	0,3616667	│	
							│		└	0,19
					┌	0,4133333	│			
					│		│		┌	0,2266667
					│		└	0,3633333	│	
					│				└	0,2066667
▶			┌	0,5383334	│					
			│		│				┌	0,1416667
			│		│		┌	0,26	│	
			│		│		│		└	0,22
			│		└	0,4133333	│			
			│				│		┌	0,05
			│				└	0,365	│	
			│						└	0,05
	┌	0,7383333	│							
	│		│						┌	0,13
	│		│				┌	0,305	│	
	│		│				│		└	0,1033333
	│		│		┌	0,4183333	│			
	│		│		│		│		┌	0,12
	│		│		│		└	0,335	│	
	│		│		│				└	0,11
	│		└	0,58	│					
	│				│				┌	0,1883333
	│				│		┌	0,2583333	│	
	│				│		│		└	0,1816667
	│				└	0,43	│			
	│						│		┌	0,02666667
	│						└	0,3516667	│	
	│								└	0,02666667
	│									
	│								┌	0,2116667
	│						┌	0,2466667	│	
	│						│		└	0,1833333
	│				┌	0,3666667	│			
	│				│		│		┌	0,1133333

3.2.11. The PedKnow Table

The PedKnow (PedKnow and PedKnowRef) Tables are generated by default for the whole population by using the Pedigree Content submenu. The PedKnow table gives the completeness (Complet) of the pedigree (MacCluer et al., 1983) for each extant parental generation in the pedigree (fathers = Generation 1; grandfathers = Generation 2; etc.). The PedKnowRef table is generated by clicking on the button asking the user for a predefined reference population.

Figure 13: Pedknow Table obtained from the example input file

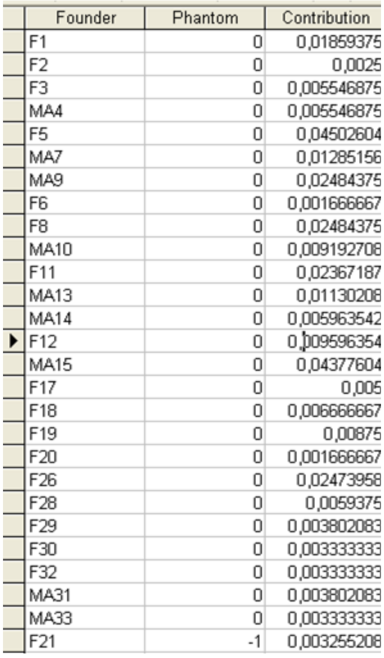


Generation	Complet
1	0,7491667
2	0,5429167
3	0,3914583
4	0,2553125
5	0,1221875
6	0,03861979
7	0,006393229
8	0,000546875
9	6,510417E-05

3.2.12. The Founders Table

The Founder Tables (Founders and FoundersRef) list the founders in the analysed population and their contribution (AR) to the population. When one parent of a listed animal is unknown its contribution to the population is that corresponding to the 'Phantom' founder. This case is identified with a Boolean field (True '1' if only one parent is known) named Phantom. The FoudersRef table is generated by clicking on the button asking the user for a predefined reference population.

Figure 14: Founders table obtained from the Gener.mdb results file



Founder	Phantom	Contribution
F1	0	0,01859375
F2	0	0,0025
F3	0	0,005546875
MA4	0	0,005546875
F5	0	0,04502604
MA7	0	0,01285156
MA9	0	0,02484375
F6	0	0,001666667
F8	0	0,02484375
MA10	0	0,009192708
F11	0	0,02367187
MA13	0	0,01130208
MA14	0	0,005963542
F12	0	0,009596354
MA15	0	0,04377604
F17	0	0,005
F18	0	0,006666667
F19	0	0,00875
F20	0	0,001666667
F26	0	0,02473958
F28	0	0,0059375
F29	0	0,003802083
F30	0	0,003333333
F32	0	0,003333333
MA31	0	0,003802083
MA33	0	0,003333333
F21	-1	0,003255208

3.2.13. The Ancestors Tables

The Ancestors table includes the information on ancestors (founders or not) explaining the genetic variability of the population identified using Boichard et al's (1997) methodology. The fields containing the information are identified as: SEL (the order in which ancestor has been selected), FUN (the identification of the selected ancestor), MIN and MAX (the maximum number POBL (the cumulated proportion of genetic variance explained by the selected ancestors), TEMIN and TEMAX (are the minimum and maximum effective number of ancestors). When computations end TEMIN and TEMAX are the same; this value is the effective number of ancestors. The AncestorsRef table is computed by clicking on the corresponding box in order to select the field defining the individuals of a given reference population (coded as '1').

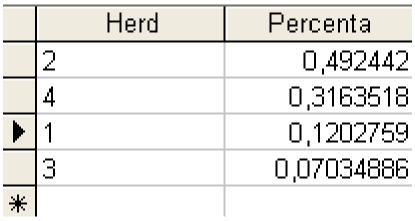
Figure 15: Ancestors table obtained from the Gener.mdb results file

SEL	FUN	MIN	MAX	POBL	TEMIN	TEMAX
1	43	1	119	0,1142691	8,94221	50,74621
2	47	2	118	0,2209977	9,505427	33,69291
3	60	3	118	0,2848028	13,60924	30,333
4	168	4	116	0,3242459	17,68331	29,28048
5	70	5	117	0,3613689	18,20227	28,49417
6	93	6	116	0,3957367	18,82728	27,81337
7	76	7	117	0,4252465	19,90088	27,39091
8	38	8	118	0,4520737	20,51343	27,06041
9	303	9	116	0,4781758	20,60944	26,70139
10	59	10	116	0,5039153	20,74416	26,39177
11	104	11	117	0,525377	21,6596	26,20911
12	85	12	116	0,5442286	22,25926	26,0537
13	78	13	115	0,56279	22,34159	25,90442
14	71	14	115	0,5804814	22,5224	25,78321
15	119	15	113	0,5978828	22,56656	25,64419
16	171	16	111	0,6147042	22,68341	25,51617
17	96	17	111	0,6305104	22,90446	25,42564
18	55	18	111	0,6460267	22,9492	25,33824
19	91	19	112	0,6599478	23,25633	25,28065
20	208	20	110	0,6738689	23,25633	25,19644
21	113	21	110	0,6870649	23,37844	25,13775
22	323	22	109	0,6992459	23,55123	25,08242
23	61	23	109	0,7108468	23,63588	25,04028
24	118	24	109	0,7224478	23,63588	24,9973
25	211	25	107	0,7328886	23,82688	24,95188
26	127	26	107	0,7430394	23,8685	24,92196
27	75	27	108	0,7529002	23,90141	24,8997

3.2.14. The FoundHerd Tables

In the Founders submenu screen users will find the facility to compute the genetic representation of the herds at the whole and at the reference population levels. It is carried out simply summing up Boichard et al.'s contribution values of the ancestors belonging to each herd. The FoundHerd table contains the identification of the herds (in the Herd field) and the percentage of the genetic representation of their founders at population level (in the Percenta field). The FoundHerdRef table is generated using only the individuals included in a predefined (labelling the individuals with '1') reference population. Note that, by default, the FoundHerd table is generated accounting for the individuals in the database with both parents known. However, if the user has previously generated the Ancestors table using a predefined reference population ENDOG will automatically generate the FoundHerdRef table and the FoundHerd table will not be generated in the current session of ENDOG.

Figure 16: FoundHerd table obtained from the Gener.mdb results file



Herd	Percenta
2	0,492442
4	0,3163518
1	0,1202759
3	0,07034886

3.2.15. The Founders.txt file

This file contains the information on the base population and effective number of founders.

Figure 17: Founders.txt file obtained from the example input file

```
    Size of Population ...    600

    Base Population (one or more unknown parents)...    169.
 Actual Base Population (one unknown parent = half founder) ...
150,5

    Effective Population Size of Founders ...     67,73.
 Expected Inbreeding caused by unbalancing of founders contribution
... 0,74%.
 Computed Mean Inbreeding ... 3,63%
```

3.2.16. The Ancestor.txt file

This file contains the information on the base population and effective number of founders.

Figure 18: Ancestor.txt file obtained from the example input file

```
Reference Populations is taken as the animals with both parents known. This
population will be smaller than that one used to analyze Founders. You may
then choose a particular population
 Number of animals in the Reference Population:      431
 Number of Ancestors contributing to the Reference Population:      92
 Effective Number of Ancestors for the Reference Population:      25
 Nº of ancestors explaining 50%:  10
```

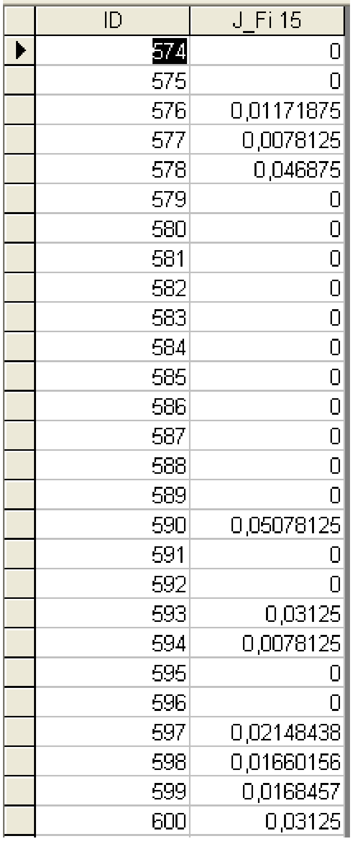
3.2.17. The F_? and F_a_? Tables

ENDOG v4.8 allows users to compute the partial inbreeding coefficients due to a given founder or ancestor for each individual a reference subpopulation by clicking on the Partial Inbreeding submenu. Clicking on the Partial Inbreeding submenu a dialog box will appear asking the user to select the column or field identifying the reference subpopulation and to choose the number (*n*) of founders and ancestors used to compute partial inbreeding coefficients. The *n* founders or ancestors selected will those *n* with the highest contributions to the population. Note that the Founders submenu must be run before running the Partial Inbreeding submenu. Additionally, ENDOG allows user choosing arbitrarily the founders or ancestors from which partial inbreeding coefficients are computed. Results will be stored in different Tables with names including the identification (the individual 15 in the Figure below) of the founder (F_?) or ancestor (F_a_?). When the founders or ancestors are chosen by the user, the name of the to which the results are written will be F_c_?, with the '?' corresponding to the identification of the founder or ancestor. Each results Table will include the identification of the individuals in the reference subpopulations (Id) and the partial inbreeding coefficient for these individuals due to a given founder or ancestor (J_F_?; see figure 16, below). If a given founder has been also identified as ancestor the corresponding results will be written only once.

3.2.18. The GenEquiv_Subpopulation Table

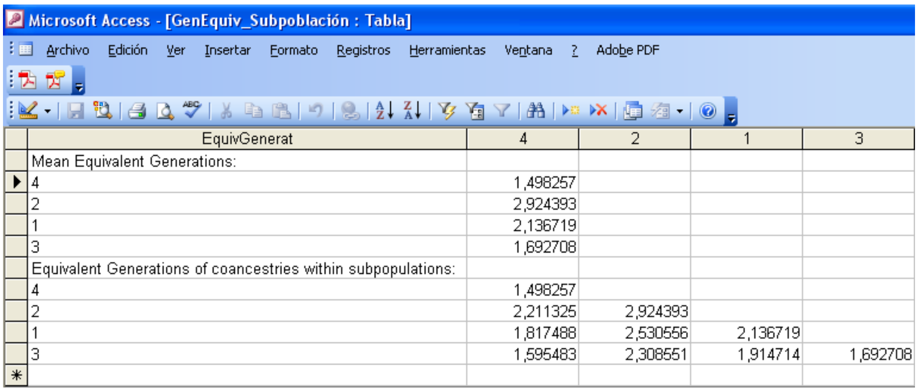
After cliking on the Subpopulations submenu and selecting the field identifying the subpopulations, a table named GenEquiv_Subpopulation is generated by default. This table contains the mean equivalent to complete generations for each subpopulation and the mean equivalent to complete generations between subpopulations. The information contained in this table may be useful to interpret the results obtained when the option to adjust coancestries by pedigree depth is used.

Figure 19: F_? table obtained from the Gener.mdb results file



ID	J_Fi 15
574	0
575	0
576	0,01171875
577	0,0078125
578	0,046875
579	0
580	0
581	0
582	0
583	0
584	0
585	0
586	0
587	0
588	0
589	0
590	0,05078125
591	0
592	0
593	0,03125
594	0,0078125
595	0
596	0
597	0,02148438
598	0,01660156
599	0,0168457
600	0,03125

Figure 20: GenEquiv_Subpopulation Table table obtained from the Gener.mdb results file



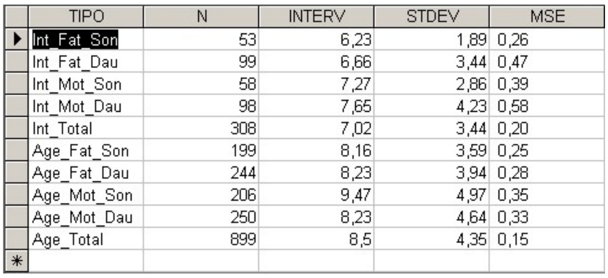
EquivGenerat	4	2	1	3
Mean Equivalent Generations:				
4	1,498257			
2	2,924393			
1	2,136719			
3	1,692708			
Equivalent Generations of coancestries within subpopulations:				
4	1,498257			
2	2,211325	2,924393		
1	1,817488	2,530556	2,136719	
3	1,595483	2,308551	1,914714	1,692708

3.2.19. The GenInterv Tables

This table contains both the generation intervals and the average age of parents at the birth of their offspring computed for the 4 pathways (father- son, father - daughter, mother - son and mother - daughter). The structure of the Table is: TIPO (the type of path for interval - Int - or average age – Age -); N (the number of data used); INTERV (average values); STDEV (the corresponding standard deviations); and MSE (the corresponding standard errors). The user can define a given subpopulation in the input data (coding the individuals a '1' in a separate column) in order to compute the generation intervals for this subpopulation. This latter result will be stored in the

GenIntRef table. Note that computation of generation intervals can be done on columns (or fields) including unknown birth dates.

Figure 21: GenInterv table obtained from the Gener.mdb results file



TIPO	N	INTERV	STDEV	MSE
Int_Fat_Son	53	6,23	1,89	0,26
Int_Fat_Dau	99	6,66	3,44	0,47
Int_Mot_Son	58	7,27	2,86	0,39
Int_Mot_Dau	98	7,65	4,23	0,58
Int_Total	308	7,02	3,44	0,20
Age_Fat_Son	199	8,16	3,59	0,25
Age_Fat_Dau	244	8,23	3,94	0,28
Age_Mot_Son	206	9,47	4,97	0,35
Age_Mot_Dau	250	8,23	4,64	0,33
Age_Total	899	8,5	4,35	0,15

3.2.20. The NeGenInterv.txt files

These files contain the N_e obtained from regression over birth date and from Log-regression on birth date for the whole population of for a given reference population.

Figure 22: NeGenInterval.txt file obtained from the example input file

```
WHOLE POPULATION.
 600 individuals
Effective size obtained from regression on the birth date:    35,71
Effective size obtained from Log regression on the birth date:    33,10
```

Figure 23: NeGenInterval_Ref.txt file obtained from the example input file

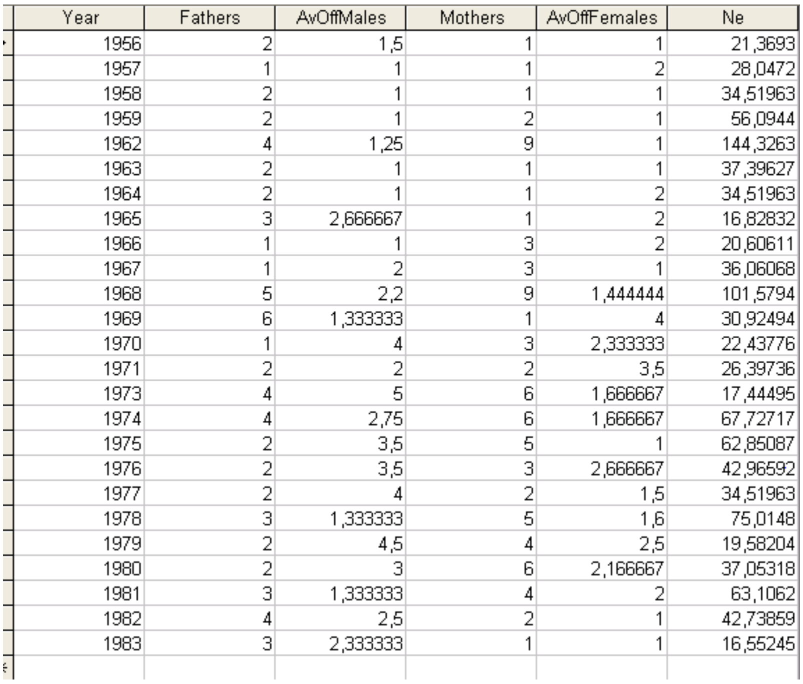
```
REFERENCE POPULATION DEFINED BY USERWHEN THE FIELD cod_alive CONTAINS 1.
 401 individuals
Effective size obtained from regression on the birth date:    32,45
Effective size obtained from Log regression on the birth date:    30,40
```

3.2.21. Ne estimates based on family size variance: the Tables NeOffs_Year, NeOffs_Gen and OffsNeRef

Tables NeOffs_Year, NeOffs_Gen and OffsNeRef are computed using the Offspring Analysis submenu. Note that for the estimation of N_e based on family size variance the program needs to know the average value of the generation interval for the analysed population. In consequence, users need to use the Generation Intervals submenu before using the Offspring Analysis submenu. In order to ascertain historical bottlenecks in the population the NeOffs_Year and NeOffs_Gen tables give estimates of N_e by the year or the period of birth of the reproductive individual respectively. The period of birth is fitted by default by rounding the average generation interval, thus approaching successive generations in the pedigree. The OffsNeRef table is computed by clicking on the corresponding box for the reproductive individuals contained in a predefined reference population. The (reproductive or not) individuals included in the reference population must be coded as 1 whilst the others can be coded using any other value. Obviously, if the reference population selected consists of all the individuals in the pedigree the N_e statistics is computed for the whole population.

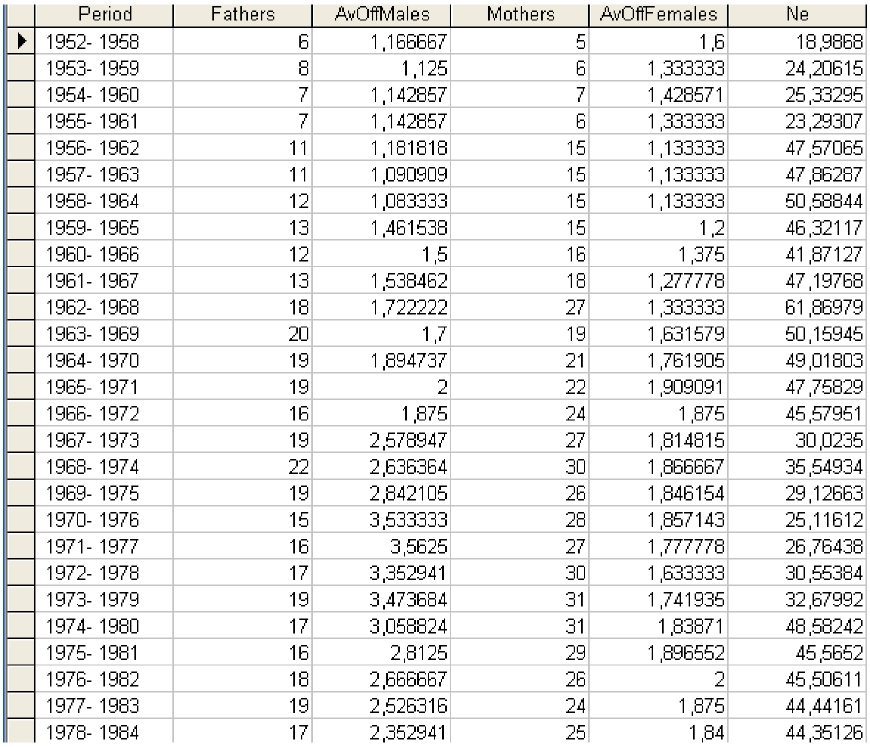
When estimates of N_e are computed on a predefined reference population, ENDOG (v4.8) assumes that the population on which the analysis is carried out corresponds to one generation. Otherwise users must carry out a transformation of the N_e estimate given by ENDOG by a factor (L/y) where L is the generation interval in years and y the number of years included in the fitted period.

Figure 24: NeOffs_Year table obtained from the Gener.mdb results file



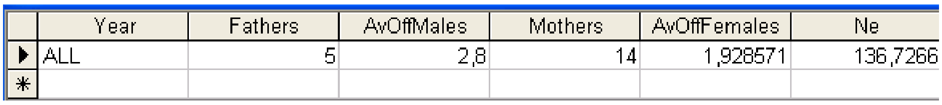
Year	Fathers	AvOffMales	Mothers	AvOffFemales	Ne
1956	2	1,5	1	1	21,3693
1957	1	1	1	2	28,0472
1958	2	1	1	1	34,51963
1959	2	1	2	1	56,0944
1962	4	1,25	9	1	144,3263
1963	2	1	1	1	37,39627
1964	2	1	1	2	34,51963
1965	3	2,666667	1	2	16,82832
1966	1	1	3	2	20,60611
1967	1	2	3	1	36,06068
1968	5	2,2	9	1,444444	101,5794
1969	6	1,333333	1	4	30,92494
1970	1	4	3	2,333333	22,43776
1971	2	2	2	3,5	26,39736
1973	4	5	6	1,666667	17,44495
1974	4	2,75	6	1,666667	67,72717
1975	2	3,5	5	1	62,85087
1976	2	3,5	3	2,666667	42,96592
1977	2	4	2	1,5	34,51963
1978	3	1,333333	5	1,6	75,0148
1979	2	4,5	4	2,5	19,58204
1980	2	3	6	2,166667	37,05318
1981	3	1,333333	4	2	63,1062
1982	4	2,5	2	1	42,73859
1983	3	2,333333	1	1	16,55245

Figure 25: NeOffs_Gen table obtained from the Gener.mdb results file



Period	Fathers	AvOffMales	Mothers	AvOffFemales	Ne
1952- 1958	6	1,166667	5	1,6	18,9868
1953- 1959	8	1,125	6	1,333333	24,20615
1954- 1960	7	1,142857	7	1,428571	25,33295
1955- 1961	7	1,142857	6	1,333333	23,29307
1956- 1962	11	1,181818	15	1,133333	47,57065
1957- 1963	11	1,090909	15	1,133333	47,86287
1958- 1964	12	1,083333	15	1,133333	50,58844
1959- 1965	13	1,461538	15	1,2	46,32117
1960- 1966	12	1,5	16	1,375	41,87127
1961- 1967	13	1,538462	18	1,277778	47,19768
1962- 1968	18	1,722222	27	1,333333	61,86979
1963- 1969	20	1,7	19	1,631579	50,15945
1964- 1970	19	1,894737	21	1,761905	49,01803
1965- 1971	19	2	22	1,909091	47,75829
1966- 1972	16	1,875	24	1,875	45,57951
1967- 1973	19	2,578947	27	1,814815	30,0235
1968- 1974	22	2,636364	30	1,866667	35,54934
1969- 1975	19	2,842105	26	1,846154	29,12663
1970- 1976	15	3,533333	28	1,857143	25,11612
1971- 1977	16	3,5625	27	1,777778	26,76438
1972- 1978	17	3,352941	30	1,633333	30,55384
1973- 1979	19	3,473684	31	1,741935	32,67992
1974- 1980	17	3,058824	31	1,83871	48,58242
1975- 1981	16	2,8125	29	1,896552	45,5652
1976- 1982	18	2,666667	26	2	45,50611
1977- 1983	19	2,526316	24	1,875	44,44161
1978- 1984	17	2,352941	25	1,84	44,35126

Figure 26: OffsNeRef table obtained from the Gener.mdb results file

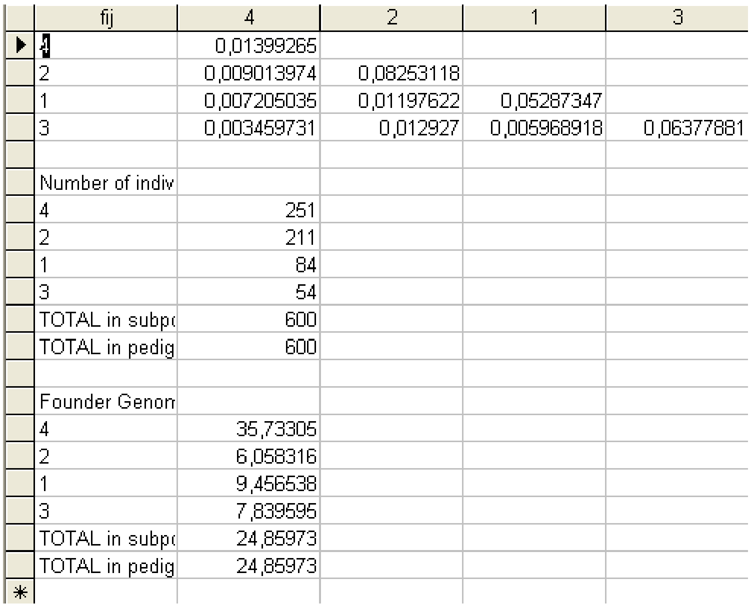


Year	Fathers	AvOffMales	Mothers	AvOffFemales	Ne
ALL	5	2,8	14	1,928571	136,7266

3.2.22. Paired distances between subpopulations: the Tables fij_?, AverDist_?, DistNei_? and Fis_Fsts_?

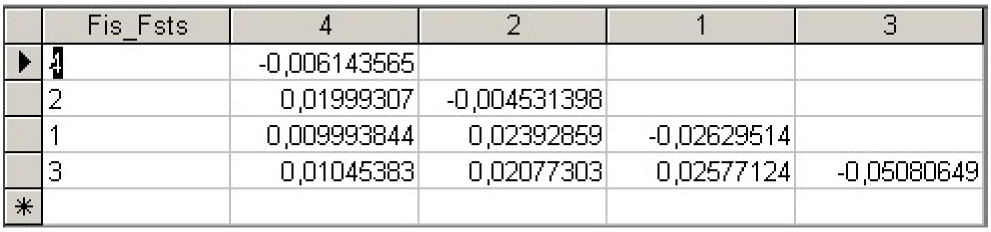
Tables fij, AverDist, DistNei and Fis_Fsts are computed using the Fstats submenu. The Table fij, is computed and written to disk by default whilst the others are computed and written on request. The table fij also stores the number of individuals and the founder genome equivalents (f_g; Ballou and Lacy, 1995) for each subpopulation. Regardless of the distance computed (Average distance, Nei's standard distance or paired Fst's) the structure of the Tables is the same: a lower left matrix containing the distance values for each pair of subpopulations. However, when the user asks ENDOG for the Fst matrix, the program will write the Fis values for each subpopulation on the diagonal. If the number of subpopulations exceeds the maximum number of columns fitted in the ACCESS tables the corresponding distance matrices are saved to disk in .txt format. Each of the tables described is named using the corresponding root (fij, AverDist, DistNei and Fis_Fsts) followed by an underscore and the name of the field in which the subpopulations are fitted (v.g. fij_Area). When the analyses are performed using the 'Fit to Equal Populations Size' option the name will be ended with an underscore and the size to which the analysis is fitted (v.g. fij_Area_150).

Figure 27: fij table obtained from the Gener.mdb results file



fij	4	2	1	3
4	0,01399265			
2	0,009013974	0,08253118		
1	0,007205035	0,01197622	0,05287347	
3	0,003459731	0,012927	0,005968918	0,06377881
Number of indiv				
4	251			
2	211			
1	84			
3	54			
TOTAL in subp	600			
TOTAL in pedig	600			
Founder Genon				
4	35,73305			
2	6,058316			
1	9,456538			
3	7,839595			
TOTAL in subp	24,85973			
TOTAL in pedig	24,85973			

Figure 28: Fis_Fsts table obtained from the Gener.mdb results file



Fis_Fsts	4	2	1	3
4	-0,006143565			
2	0,01999307	-0,004531398		
1	0,009993844	0,02392859	-0,02629514	
3	0,01045383	0,02077303	0,02577124	-0,05080649

3.2.23. The MatFs_?t.txt file

This file complements the Fis_Fsts Table above and contains the values of the F-statistics (Fis, Fit and Fst) for the whole population.

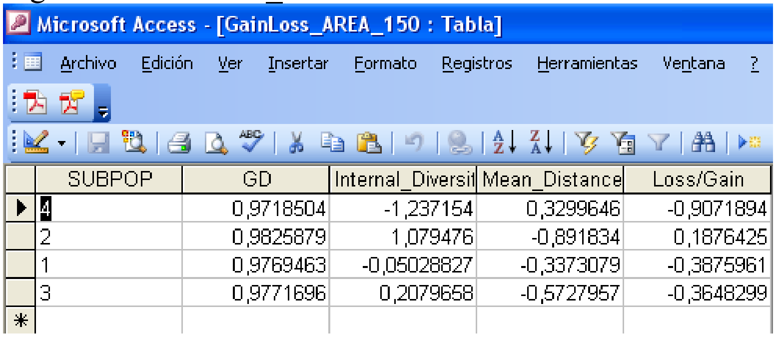
Figure 29: MatFst_?.txt file obtained from the example input file

```
Subpopulations:= 4

POPULATION AVERAGE OF:
Mean Coancestry within Subpopulations:=  0,048002
Selfcoancestry:=  0,518128
Inbreeding:=  0,036256
Nei Distance:=  0,027896
Mean Coancestry in the Metapopulation:=  0,020106

Wright F Parameters:
Fis=- 0,012339
Fst= 0,028468
Fit= 0,016481
```

3.2.24. The GainLoss_? Tables

After performed a population structure analysis using the Population Submenu, user can click on the Gain/Loss box. After doing it, the genetic contributions for each subpopulation to total diversity will be computed. The table generated will inlcude the identification of the subpopulations (SUBPOP), the Nei's (1987) Gene Diversity remaining after removal of the corresponding supopulation (GD) and the contributions (in percentage) of each subpopulation to within-subpopulations (Internal_Diversity), between-subpopulations (Mean_Distance) and total (Loss/Gain) gene diversity. The GainLoss tables are named continuing the root with an underscore and the name of the field in which the subpopulations are fitted (v.g. GainLoss_Area). When the analyses is performed using the 'Fit to Equal Populations Size' option the name of the table will be ended with an underscore and the size to which the analysis is fitted (v.g. GainLoss_Area_150; see the figure below). User can select for conservation an arbitarry number of subpopulations using the "Add to list" and "Remove from list" boxes, The results obtained will be stored in the corresponding tables below the general results.

Figure 30: GainLoss_? table obtained from the Gener.mdb results file

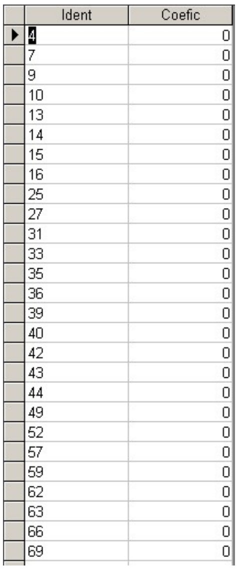


SUBPOP	GD	Internal_Diversit	Mean_Distance	Loss/Gain
4	0,9718504	-1,237154	0,3299646	-0,9071894
2	0,9825879	1,079476	-0,891834	0,1876425
1	0,9769463	-0,05028827	-0,3373079	-0,3875961
3	0,9771696	0,2079658	-0,5727957	-0,3648299

3.2.25. The Parent Table

As explained in the corresponding section this table keeps the information on all possible matings we can carry out using a key individual. In this table there are two fields: a) Ident field which is the identification of all the possible individuals to mate with the key individual and b) Coefic field showing the corresponding coancestry coefficients with the key individual.

Figure 31: Parent table obtained from the Gener.mdb results file

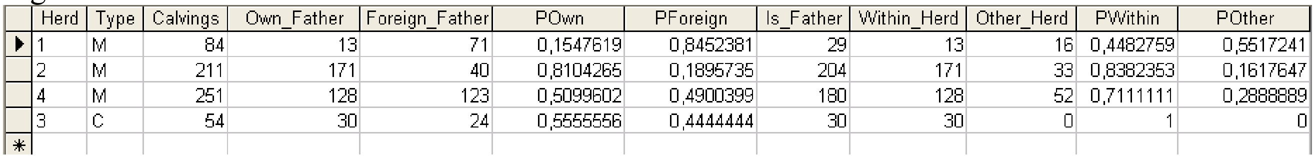


Ident	Coefic
4	0
7	0
9	0
10	0
13	0
14	0
15	0
16	0
25	0
27	0
31	0
33	0
35	0
36	0
39	0
40	0
42	0
43	0
44	0
49	0
52	0
57	0
59	0
62	0
63	0
66	0
69	0

3.2.26. The HerStr Table

This Table contains, for each individual herd, the information that will be further summarized in the StrHer Table (see next section). The fields containing the information are identified as: Herd (identification of each individual herd); Type (the classification obtained for each individual herd as nucleus, multiplier, commercial or disconnected); Calvings (number of registered individuals born in the herd); Own_Father (number of individuals born in the herd with father born in the same herd); Foreign_Father (number of individuals born in the herd with father born in other herd); Pown (ratio Own_Father / Calvings); Pforeign (ratio Foreign_Father / Calvings); Is_Father (number of times in which the reproductive males born in the herd act as fathers); Within_Herd (number of times in which the reproductive males born in the herd act as fathers in the same herd); Other_Herd (number of times in which the reproductive males born in the herd act as fathers in other herd); PWithin (ratio Within_Herd / Is_Father in percentage); POther (ratio Other_Herd / Is_Father in percentage).

Figure 32: HerStr table obtained from the Gener.mdb results file

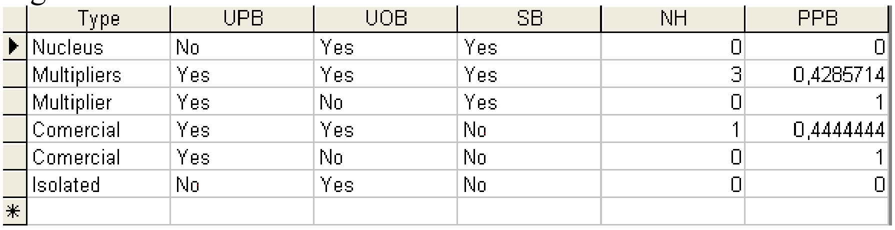


Herd	Type	Calvings	Own_Father	Foreign_Father	POwn	PForeign	Is_Father	Within_Herd	Other_Herd	PWithin	POther
1	M	84	13	71	0,1547619	0,8452381	29	13	16	0,4482759	0,5517241
2	M	211	171	40	0,8104265	0,1895735	204	171	33	0,8382353	0,1617647
4	M	251	128	123	0,5099602	0,4900399	180	128	52	0,7111111	0,2888889
3	C	54	30	24	0,5555556	0,4444444	30	30	0	1	0

3.2.27. The StrHer Table

This Table summarizes the information detailed in the HerStr Table above. Names of fields included in this table are: Type (if the herds are considered nucleus, multipliers, commercial or disconnected), UPB (use of purchased reproductive males), UOB (use their own reproductive males), SB (sell males used for reproduction), NH (number of herds in each class), PPB (percent of purchased reproductive males).

Figure 33: StrHerd table obtained from the Gener.mdb results file

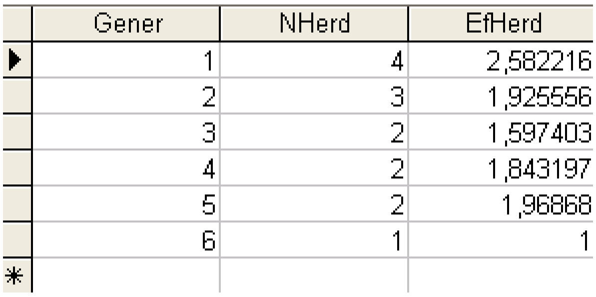


Type	UPB	UOB	SB	NH	PPB
Nucleus	No	Yes	Yes	0	0
Multipliers	Yes	Yes	Yes	3	0,4285714
Multiplier	Yes	No	Yes	0	1
Comercial	Yes	Yes	No	1	0,4444444
Comercial	Yes	No	No	0	1
Isolated	No	Yes	No	0	0

3.2.28. The Robert Table

This table gives the information on the actual and effective number of herds supplying male ancestors. The generation in which the male ancestor is found is shown in the Gener field within the table (1 for fathers, 2 for grandfathers, 3 for great-grandfathers and so on). NHerd and EfHerd show, respectively, the actual and the effective number of herds in each generation.

Figure 34: Robert table obtained from the Gener.mdb results file



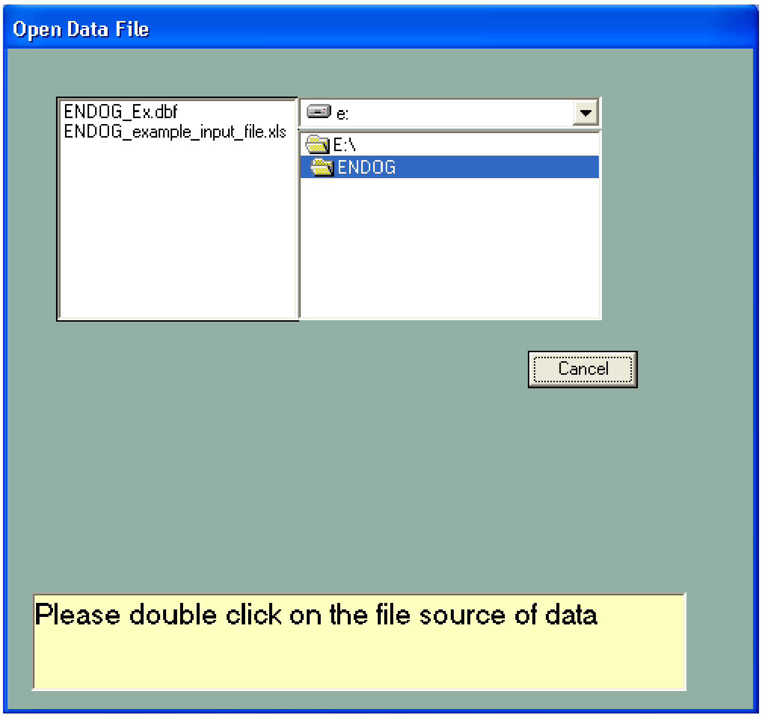
Gener	NHerd	EfHerd
1	4	2,582216
2	3	1,925556
3	2	1,597403
4	2	1,843197
5	2	1,96868
6	1	1

3.3 A Session with ENDOG (v 4.8)

As pointed out in the section 3.1, at the beginning of a session ENDOG will ask the user if individuals contained in the data are sequentially ordered. If the answer is NO, the program will ask for the column (or field) identifying birth date. Whatever the answer, ENDOG will check consistency of data and, if errors are found, will write a text file (error.txt) including those animals with errors in sex or birth date. If not all the records have actual birth dates the data set must be previously ordered. If the user does not wish to reorder the dataset virtual birth dates must be created for the individuals lacking this information (in the same or in a different column or field as that containing actual birth dates).

After the user has resolved any problems with the input data, ENDOG will compute the individual inbreeding and average relatedness coefficient as well as the number of full generations traced, the maximum number of generations traced, the equivalent complete generations, the offspring size and the individual increase in inbreeding for each individual. In addition, the main screen of ENDOG has four boxes allowing users to compute the Inbred_? Table (that on the left), the Coancest.txt file and the Coan_Ref.txt file and the GCI table.

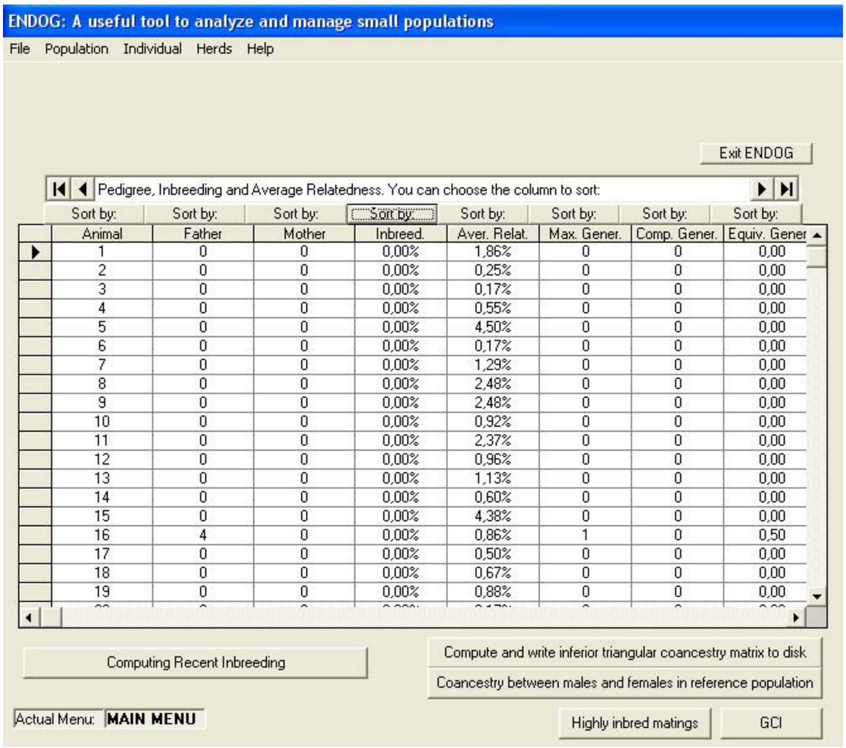
Figure 35: Initial screen of ENDOG



Please double click on the file source of data

The user will find three different menus: Population, Individuals and Herds. The Population Menu has 7 different submenus: Inbreeding per Generation, Pedigree Content, Founders, Partial Inbreeding, Generation Intervals, Offspring Analysis and Subpopulations (Fstats).

Figure 36: ENDOG Main Screen



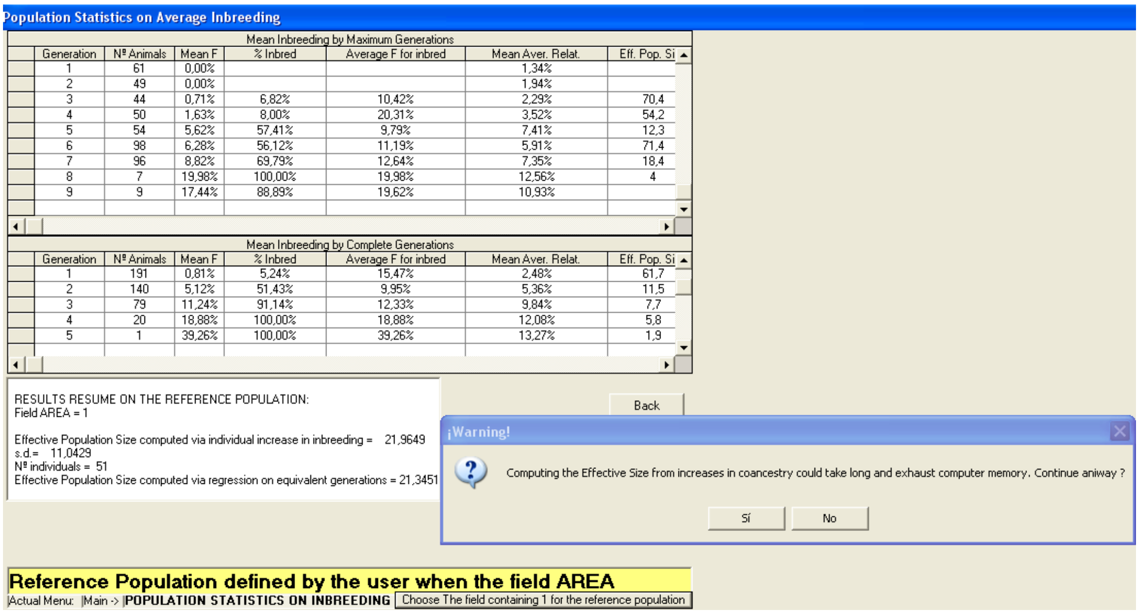
Pedigree, Inbreeding and Average Relatedness. You can choose the column to sort:

Animal	Father	Mother	Inbreed.	Aver. Relat.	Max. Gener.	Comp. Gener.	Equiv. Gener
1	0	0	0,00%	1,86%	0	0	0,00
2	0	0	0,00%	0,25%	0	0	0,00
3	0	0	0,00%	0,17%	0	0	0,00
4	0	0	0,00%	0,55%	0	0	0,00
5	0	0	0,00%	4,50%	0	0	0,00
6	0	0	0,00%	0,17%	0	0	0,00
7	0	0	0,00%	1,29%	0	0	0,00
8	0	0	0,00%	2,48%	0	0	0,00
9	0	0	0,00%	2,48%	0	0	0,00
10	0	0	0,00%	0,92%	0	0	0,00
11	0	0	0,00%	2,37%	0	0	0,00
12	0	0	0,00%	0,96%	0	0	0,00
13	0	0	0,00%	1,13%	0	0	0,00
14	0	0	0,00%	0,60%	0	0	0,00
15	0	0	0,00%	4,38%	0	0	0,00
16	4	0	0,00%	0,86%	1	0	0,50
17	0	0	0,00%	0,50%	0	0	0,00
18	0	0	0,00%	0,67%	0	0	0,00
19	0	0	0,00%	0,88%	0	0	0,00

Actual Menu: MAIN MENU

The Inbreeding per Generation submenu calculates the default computations (individual figures for F, AR, generations traced and the offspring size) and N_e by number of full generations traced and maximum number of generations traced. The user can obtain these parameters for a given reference subpopulation by clicking on the box on the lower left of the screen. Moreover, clicking on this box to choose a reference population, N_e will be estimated via individual increase in inbreding and via regression on equivalente generations. At the same time, ENDOG will ask user on estimating N_e via paired increase in coancestry. NOTE that this can take a long time even if the pedigree is shallow!!! If your answer is “Si” (YES) enjoy the coffee!!!

Figure 37: Inbreeding per Generation submenu Screen



Mean Inbreeding by Maximum Generations

Generation	Nº Animals	Mean F	% Inbred	Average F for inbred	Mean Aver. Relat.	Eff. Pop. Si
1	61	0,00%			1,34%	
2	49	0,00%			1,94%	
3	44	0,71%	6,82%	10,42%	2,29%	70,4
4	50	1,63%	8,00%	20,31%	3,52%	54,2
5	54	5,62%	57,41%	9,79%	7,41%	12,3
6	98	6,28%	56,12%	11,19%	5,91%	71,4
7	96	8,82%	69,79%	12,64%	7,35%	18,4
8	7	19,98%	100,00%	19,98%	12,56%	4
9	9	17,44%	88,89%	19,62%	10,93%	

Mean Inbreeding by Complete Generations

Generation	Nº Animals	Mean F	% Inbred	Average F for inbred	Mean Aver. Relat.	Eff. Pop. Si
1	191	0,81%	5,24%	15,47%	2,48%	61,7
2	140	5,12%	51,43%	9,95%	5,36%	11,5
3	79	11,24%	91,14%	12,33%	9,84%	7,7
4	20	18,88%	100,00%	18,88%	12,08%	5,8
5	1	39,26%	100,00%	39,26%	13,27%	1,9

RESULTS RESUME ON THE REFERENCE POPULATION:
Field AREA = 1

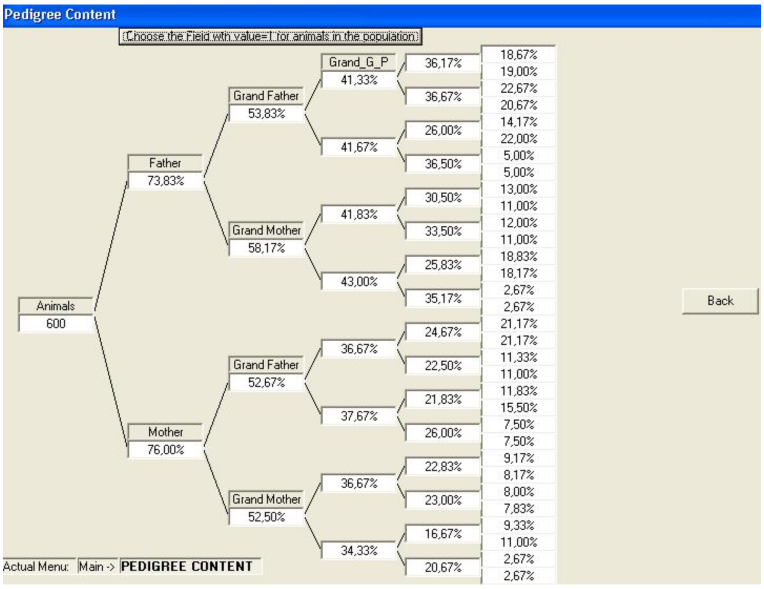
Effective Population Size computed via individual increase in inbreeding = 21,9649
s.d.= 11,0429
Nº individuals = 51
Effective Population Size computed via regression on equivalent generations = 21,3451

¡Warning!

Computing the Effective Size from increases in coancestry could take long and exhaust computer memory. Continue aniway ?

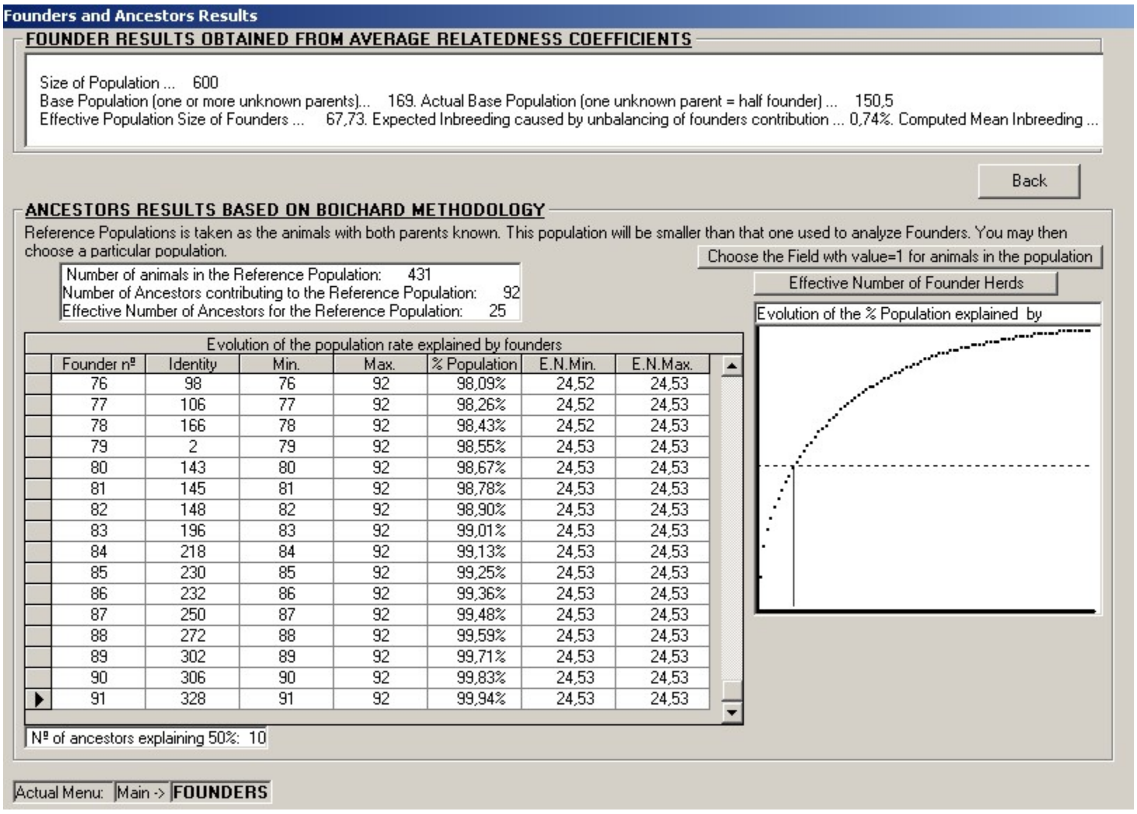
Reference Population defined by the user when the field AREA

Figure 38: Pedigree Content submenu Screen



The Pedigree Content submenu computes the completeness of the whole analysed dataset or the completeness of a reference population previously defined (by clicking on the corresponding box). The Founders submenu allows computing of the contribution of the founders to the population, the effective size of founder population, the effective number of ancestors (Boichard et al., 1997) and the effective number of founder herds. In addition, a results file with statistics for each selected ancestor or founder herd is written. To compute effective number of ancestors and herds, ENDOG asks first for a column (or field) in the input data file including the individuals forming the reference population (they must be coded as 1 and the others as any other value) and second, for the column containing the identification of the herds. Regardless of the definition of a particular reference population to compute Boichard et al's (1997) statistics, ENDOG will, by default, compute them using all the individuals with both parents known.

Figure 39: ENDOG Founders submenu Screen



Size of Population ... 600
Base Population (one or more unknown parents)... 169. Actual Base Population (one unknown parent = half founder) ... 150,5
Effective Population Size of Founders ... 67,73. Expected Inbreeding caused by unbalancing of founders contribution ... 0,74%. Computed Mean Inbreeding ...

Reference Populations is taken as the animals with both parents known. This population will be smaller than that one used to analyze Founders. You may then choose a particular population.

Number of animals in the Reference Population: 431
Number of Ancestors contributing to the Reference Population: 92
Effective Number of Ancestors for the Reference Population: 25

Evolution of the population rate explained by founders

Founder nº	Identity	Min.	Max.	% Population	E.N.Min.	E.N.Max.
76	98	76	92	98,09%	24,52	24,53
77	106	77	92	98,26%	24,52	24,53
78	166	78	92	98,43%	24,52	24,53
79	2	79	92	98,55%	24,53	24,53
80	143	80	92	98,67%	24,53	24,53
81	145	81	92	98,78%	24,53	24,53
82	148	82	92	98,90%	24,53	24,53
83	196	83	92	99,01%	24,53	24,53
84	218	84	92	99,13%	24,53	24,53
85	230	85	92	99,25%	24,53	24,53
86	232	86	92	99,36%	24,53	24,53
87	250	87	92	99,48%	24,53	24,53
88	272	88	92	99,59%	24,53	24,53
89	302	89	92	99,71%	24,53	24,53
90	306	90	92	99,83%	24,53	24,53
91	328	91	92	99,94%	24,53	24,53

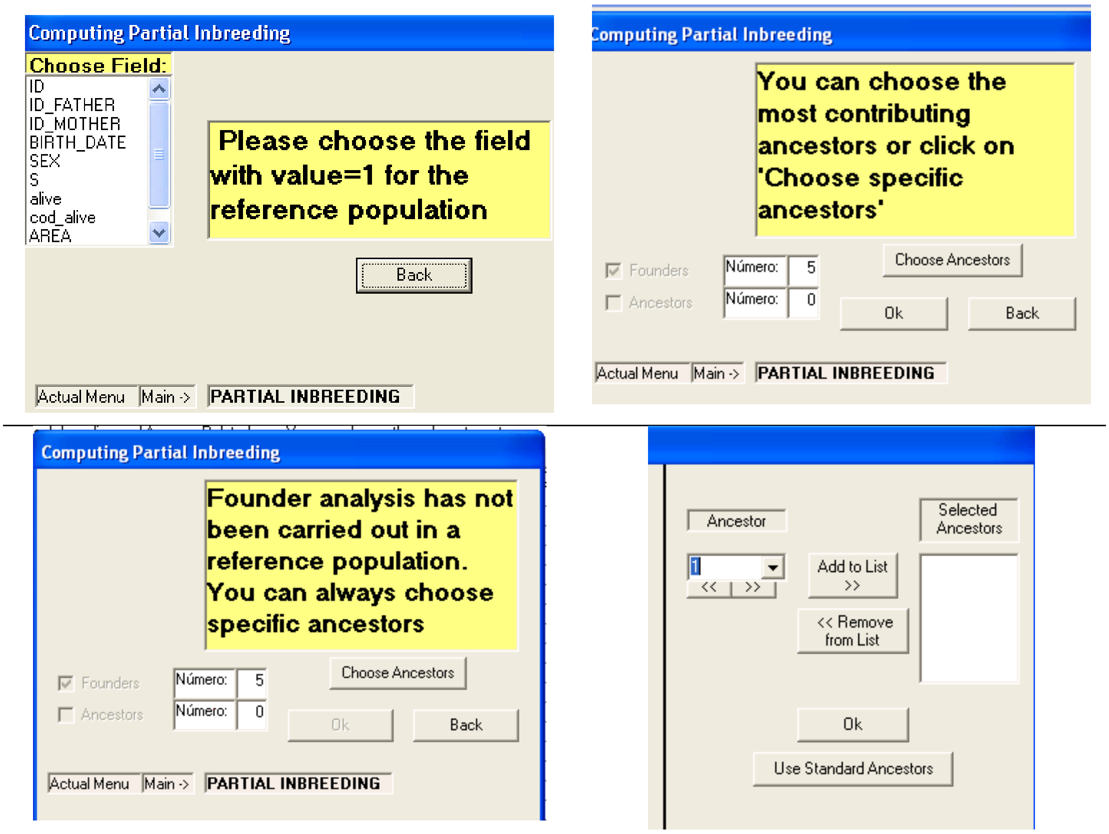
Nº of ancestors explaining 50%: 10

The Partial Inbreeding submenu allows the user to compute Lacy et al.'s (1996) partial inbreeding coefficients for a given number of founders and ancestors. The Generation Intervals submenu computes both the generation lengths and the average age (and standard errors) of parents at the birth of their offspring (kept for reproduction or not) for the 4 pathways and for the whole population.

The Offspring Analysis submenu allows users to obtain estimates of N_e according to the family size variance per period of time or predefining a given reference population by clicking on the corresponding box. Note that the period of time on which N_e is computed approaches (in years) the generation interval; so, this feature of ENDOG can not be used before computing generation intervals.

Finally, the Fstats submenu computes first the entire relationship matrix for the analyzed population and, using this, the average values for inbreeding, coancestry, Fis, Fit and Fst (Caballero and Toro, 2000; 2002) for the whole population and paired Fst, Nei's D_A and average distance values between individuals of each pair of subpopulation. Before computing the latter statistics ENDOG asks the user for a column (or field) in the input data file classifying animals within subpopulations. The values for the whole population are computed by default. However, the paired distance matrices will be saved in the corresponding ACCESS table after the user clicks on the corresponding button. Using the Fstats submenu user can also compute the contributions to total diversity of each supopulation (clicking on the “Loss/Gain” box) and adjust the desired computation by subpopulation size (clicking on the “Fit to Equal Size” box or, after clicking on the “Loss/Gain” box, clicking on the “Fitting to common size” box). See Figure 43 for a full sequence of computations when subpopulations are defined by “Area”.

Figure 40: ENDOG Partial Inbreeding submenu Screen



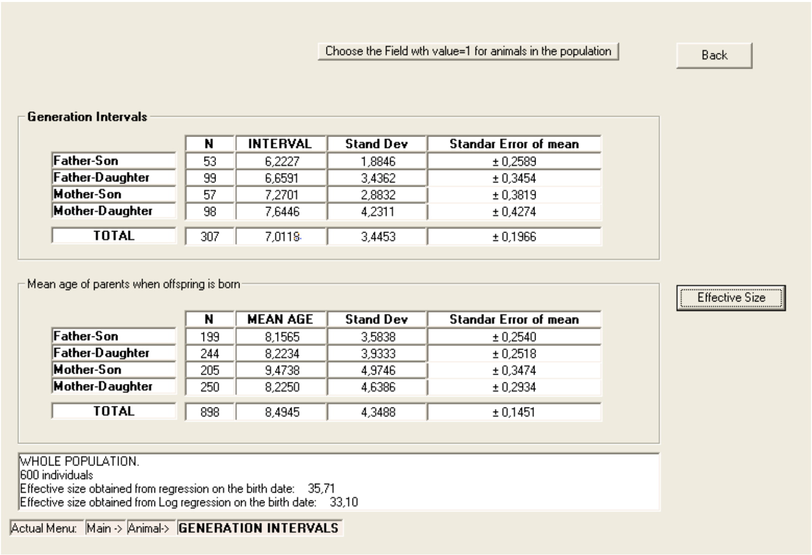
Please choose the field with value=1 for the reference population

You can choose the most contributing ancestors or click on 'Choose specific ancestors'

☑ Founders Número: 5
☐ Ancestors Número: 0

Founder analysis has not been carried out in a reference population. You can always choose specific ancestors

Figure 41: ENDOG Generation Intervals submenu Screen



Generation Intervals

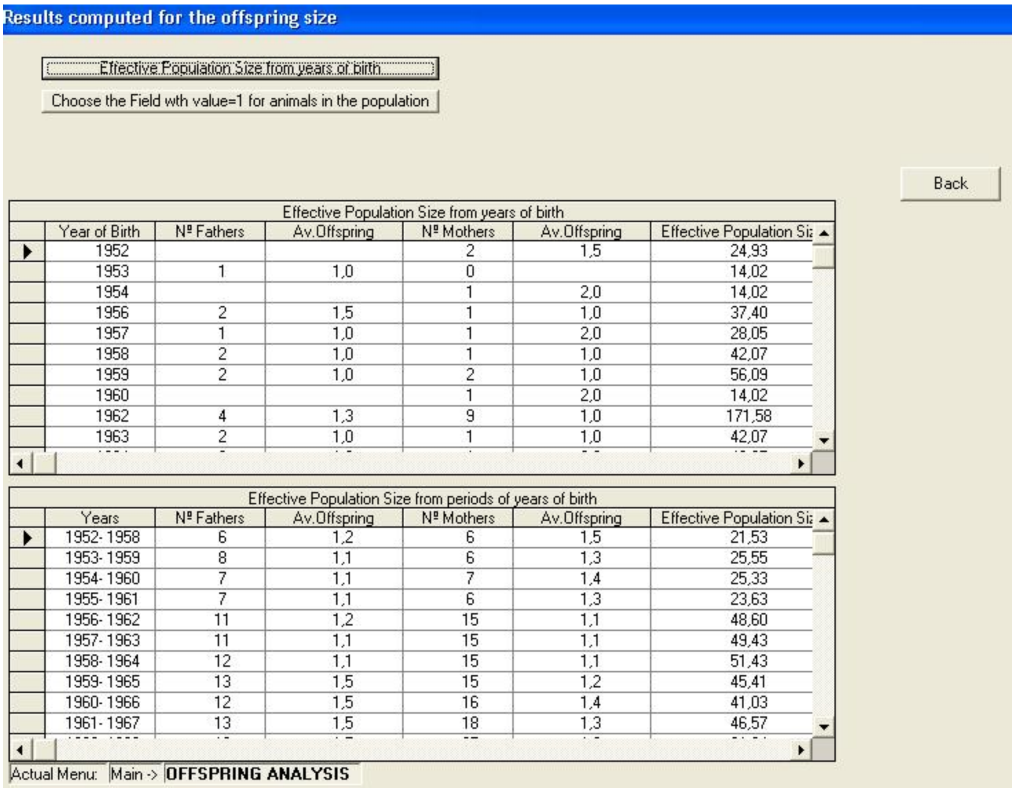
	N	INTERVAL	Stand Dev	Standar Error of mean
Father-Son	53	6,2227	1,8846	± 0,2589
Father-Daughter	99	6,6591	3,4362	± 0,3454
Mother-Son	57	7,2701	2,8832	± 0,3819
Mother-Daughter	98	7,6446	4,2311	± 0,4274
TOTAL	307	7,0118	3,4453	± 0,1966

Mean age of parents when offspring is born

	N	MEAN AGE	Stand Dev	Standar Error of mean
Father-Son	199	8,1565	3,5838	± 0,2540
Father-Daughter	244	8,2234	3,9333	± 0,2518
Mother-Son	205	9,4738	4,9746	± 0,3474
Mother-Daughter	250	8,2250	4,6386	± 0,2934
TOTAL	898	8,4945	4,3488	± 0,1451

```
WHOLE POPULATION.
600 individuals
Effective size obtained from regression on the birth date:    35,71
Effective size obtained from Log regression on the birth date:    33,10
```

Figure 42: ENDOG Offspring Analysis submenu Screen



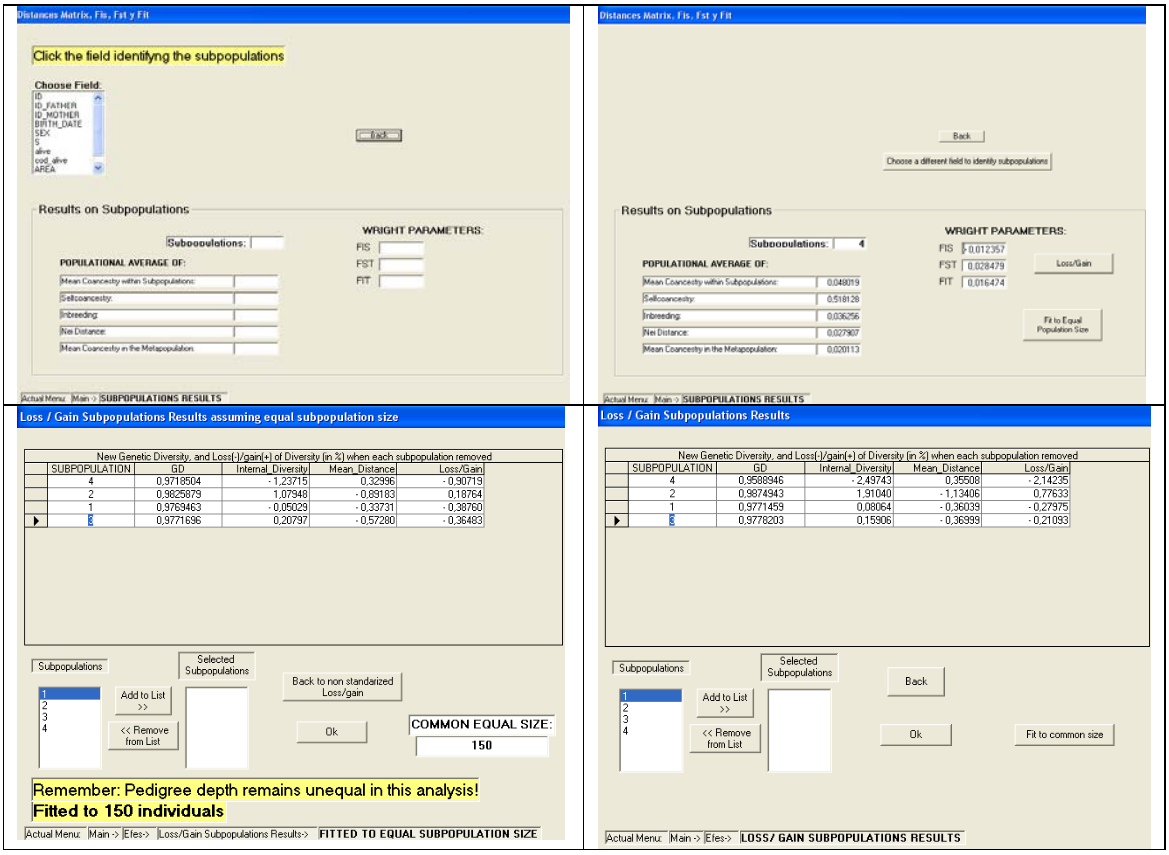
Effective Population Size from years of birth

Year of Birth	Nº Fathers	Av.Offspring	Nº Mothers	Av.Offspring	Effective Population Siz
1952			2	1,5	24,93
1953	1	1,0	0		14,02
1954			1	2,0	14,02
1956	2	1,5	1	1,0	37,40
1957	1	1,0	1	2,0	28,05
1958	2	1,0	1	1,0	42,07
1959	2	1,0	2	1,0	56,09
1960			1	2,0	14,02
1962	4	1,3	9	1,0	171,58
1963	2	1,0	1	1,0	42,07

Effective Population Size from periods of years of birth

Years	Nº Fathers	Av.Offspring	Nº Mothers	Av.Offspring	Effective Population Siz
1952- 1958	6	1,2	6	1,5	21,53
1953- 1959	8	1,1	6	1,3	25,55
1954- 1960	7	1,1	7	1,4	25,33
1955- 1961	7	1,1	6	1,3	23,63
1956- 1962	11	1,2	15	1,1	48,60
1957- 1963	11	1,1	15	1,1	49,43
1958- 1964	12	1,1	15	1,1	51,43
1959- 1965	13	1,5	15	1,2	45,41
1960- 1966	12	1,5	16	1,4	41,03
1961- 1967	13	1,5	18	1,3	46,57

Figure 43: ENDOG for Fstats submenu Screens. Sequence starts from the upper left image and continues clockwise.



Subpopulations: 4

POPULATIONAL AVERAGE OF:

Mean Coancestry within Subpopulations:	0,048019
Selfcoancestry:	0,518128
Inbreeding:	0,036256
Nei Distance:	0,027907
Mean Coancestry in the Metapopulation:	0,020113

WRIGHT PARAMETERS:

FIS	-0,012357
FST	0,028479
FIT	0,016474

New Genetic Diversity, and Loss(-)/gain(+) of Diversity (in %) when each subpopulation removed

SUBPOPULATION	GD	Internal_Diversity	Mean_Distance	Loss/Gain
4	0,9588946	- 2,49743	0,35508	- 2,14235
2	0,9874943	1,91040	- 1,13406	0,77633
1	0,9771459	0,08064	- 0,36039	- 0,27975
3	0,9778203	0,15906	- 0,36999	- 0,21093

New Genetic Diversity, and Loss(-)/gain(+) of Diversity (in %) when each subpopulation removed

SUBPOPULATION	GD	Internal_Diversity	Mean_Distance	Loss/Gain
4	0,9718504	- 1,23715	0,32996	- 0,90719
2	0,9825879	1,07948	- 0,89183	0,18764
1	0,9769463	- 0,05029	- 0,33731	- 0,38760
3	0,9771696	0,20797	- 0,57280	- 0,36483

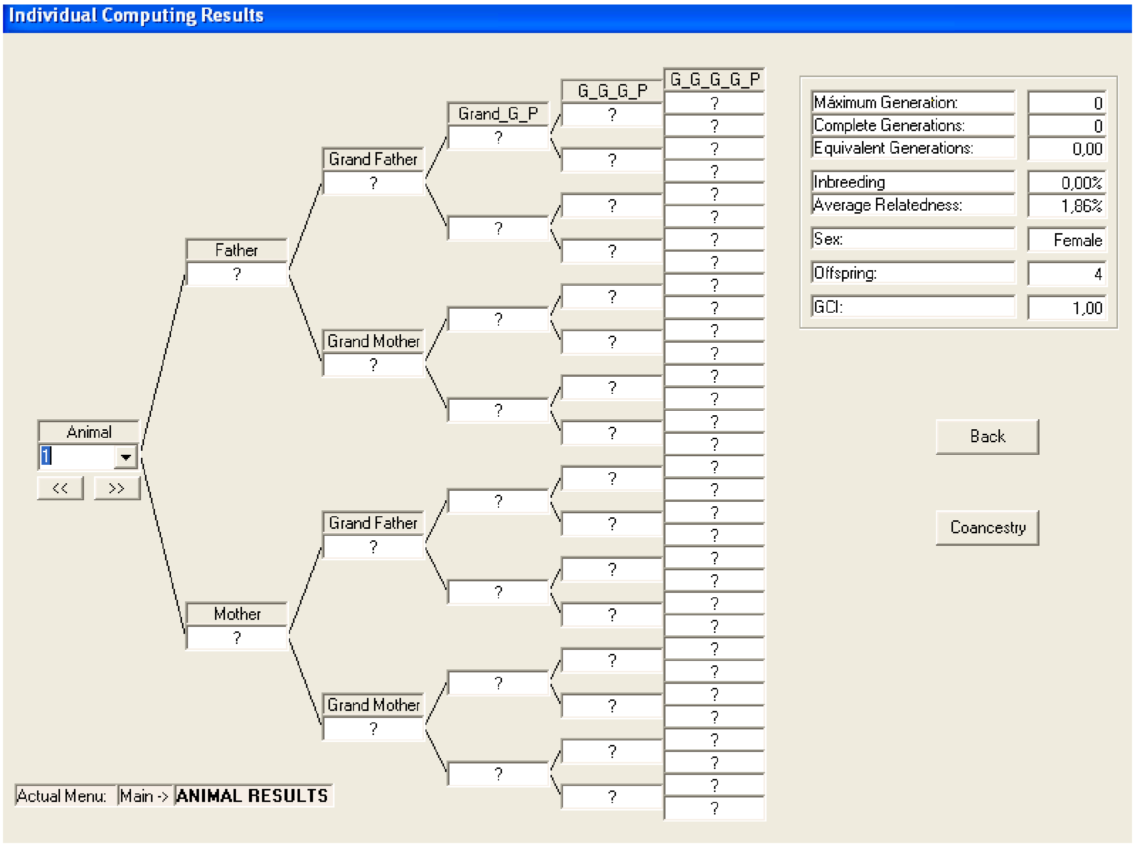
COMMON EQUAL SIZE: 150

Remember: Pedigree depth remains unequal in this analysis!

Fitted to 150 individuals

The second Menu of ENDOG is the Individuals Menu. This Menu has been provided to help teachers explain some population genetic concepts to students. It can also be of interest for breeders in the management of a given herd. The Individuals Menu has four submenus: Coancestry, Breeding Animals and Individual Pedigree. When the user clicks on the Coancestry submenu ENDOG will show all the possible individuals to be mated with the animal we previously had marked in the main screen. Alongside the individuals to be mated ENDOG shows their coancestry coefficients with the key individual. After that, the user can return to the main screen to select any other individual to calculate all their possible matings. Results of this procedure will be saved in an ACCESS table named Parent for the last individual selected. With the Breeding Individuals submenu the user can select several possible mating to calculate the average relatedness coefficient of the individuals to be mated and their coancestry coefficient.

Figure 44: ENDOG Individual Pedigree submenu Screen



Animal: 1

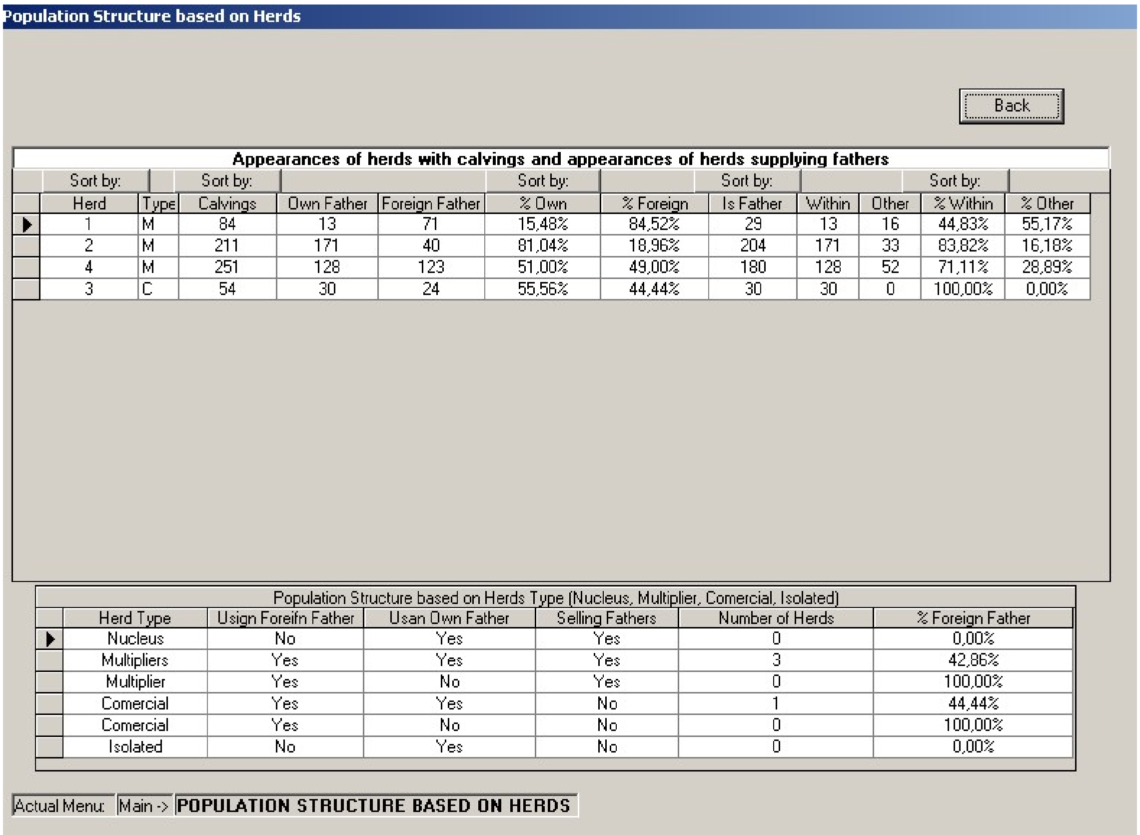
Máximum Generation: 0
Complete Generations: 0
Equivalent Generations: 0,00
Inbreeding 0,00%
Average Relatedness: 1,86%
Sex: Female
Offspring: 4
GCI: 1,00

Actual Menu: Main -> ANIMAL RESULTS

Using the Individual Pedigree submenu users can see the pedigree of any individual included in the data set and their main genealogical parameters. Clicking on the Offspring submenu, users will obtain the same results as using the Individual pedigree submenu with the addition of the identification of the offspring of the key individual.

The third Menu in ENDOG is the Herds Menu. This has two submenus: a) Population Structure by Herds; and b) Supplying Fathers, Grandfathers, etc. This descriptive allows users easy access to the main features contained there in. The former submenu (Population Structure by Herds) computes Vassallo et al.’s (1996) statistics; results are written in two different ACCESS tables including, first a summary of the statistics and then detailed statistics for each individual herd. The second submenu (Supplying Fathers, Grandfathers, etc) computes the inverse of Robertson’s (1953) probabilities that two animals taken at random in the population have their parent in the same herd for each path to know the effective number of herds supplying fathers (HS), grandfathers (HSS) and great-grandfathers (HSSS).

Figure 44: ENDOG Population Structure by Herds submenu Screen



Appearances of herds with calvings and appearances of herds supplying fathers

Herd	Type	Calvings	Own Father	Foreign Father	% Own	% Foreign	Is Father	Within	Other	% Within	% Other
1	M	84	13	71	15,48%	84,52%	29	13	16	44,83%	55,17%
2	M	211	171	40	81,04%	18,96%	204	171	33	83,82%	16,18%
4	M	251	128	123	51,00%	49,00%	180	128	52	71,11%	28,89%
3	C	54	30	24	55,56%	44,44%	30	30	0	100,00%	0,00%

Population Structure based on Herds Type (Nucleus, Multiplier, Comercial, Isolated)

Herd Type	Usign Foreifn Father	Usan Own Father	Selling Fathers	Number of Herds	% Foreign Father
Nucleus	No	Yes	Yes	0	0,00%
Multipliers	Yes	Yes	Yes	3	42,86%
Multiplier	Yes	No	Yes	0	100,00%
Comercial	Yes	Yes	No	1	44,44%
Comercial	Yes	No	No	0	100,00%
Isolated	No	Yes	No	0	0,00%

4.- References

Alderson L, 1992. A system to maximize the maintenance of genetic variability in small populations. In Genetic Conservation of Domestic Livestock II, Eds. Alderson, L. and Bodo, I., CABI, Wallingford, U.K., pp. 18-29.

Ballou JD, Lacy RC, 1995. Identifying genetically important individuals for management of genetic variation in pedigreed populations. In: Ballou JD, Gilpin M, Foose TJ, editors. Population management for survival and recovery: analytical methods and strategies in small population management. New York: Columbia University Press, pp 76-111.

Bartolomé, E., Goyache, F., Molina, A., Cervantes, I., Valera, M., Gutiérrez, J.P. (2010) Pedigree estimation of the (sub) population contribution to the total gene diversity: the horse coat colour case. *Animal*, 4, 867-875.

Boichard D, 2002. PEDIG: a FORTRAN package for pedigree analysis suited for large populations, in: Proc. 7th World Cong. Genet. Appl. to Livest. Prod., Montpellier, 19_23 August 2002, INRA, Castanet-Tolosan, France, CD-Rom, comm. No. 28-13.

Boichard D, Maignel L, Verrier E, 1997. The value of using probabilities of gene origin to measure genetic variability in a population. Genet Sel Evol 29: 5-23.

Caballero A, Toro MA, 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet Res Camb 75: 331-343.

Caballero A, Toro MA, 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conserv Gen 3: 289-299.

Cervantes, I., Goyache, F., Molina, A., Valera, M., Gutiérrez, J.P. (2008) Application of individual increase in inbreeding to estimate effective sizes from real pedigrees. J. Anim. Breed. Genet., 125, 301-310.

Cervantes, I., Goyache, F., Molina, A., Valera, M., Gutiérrez, J.P. (2010) Estimation of effective population size from the rate of coancestry in pedigreed populations. *Journal of Animal Breeding and Genetics*, in press.

Eding H, Meuwissen THE, 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. Journal of Animal Breeding and Genetics 118:141-59

Falconer DS, Mackay TFC, 1996. Introduction to Quantitative Genetics, Longman, Harlow.

González-Recio O, López de Maturana E, Gutiérrez JP, 2007. Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. J. Dairy Sci., 90, 5744-5752.

Goyache F, Gutiérrez JP, Fernández I, Gómez E, Álvarez I, Díez J, Royo LJ, 2003. Using pedigree information to monitor genetic variability of endangered populations: the Xalda sheep breed of Asturias as an example. J Anim Breed Genet 120: 95-103.

Gutiérrez JP, Altarriba J, Díaz C, Quintanilla AR, Cañón J, Piedrafita J, 2003. Genetic analysis of eight Spanish beef cattle breeds. Genet Sel Evol 35:43-64.

Gutiérrez JP, Cervantes I, Molina A, Valera M, Goyache F, 2008. Individual increase in inbreeding allows estimating realised effective sizes from pedigrees. Genet. Sel. Evol., 40, 359-378.

Gutiérrez, J.P., Cervantes, I., Goyache, F. 2009. Improving the estimation of realised effective population sizes in farm animals. J. Anim. Breed. Genet., in press.

Hill WG, 1979. A note on effective population size with overlapping generations. Genetics, 92: 317-322.

James J, 1972. Computation of genetic contributions from pedigrees. Theor. Appl. Genet. 42: 272-273.

James JW, 1977. A note on selection differentials and generation length when generations overlap. Anim. Prod. 24:109-112.

Lacy RC, 1989. Analysis of Founder Representation in Pedigrees: Founder Equivalent and Founder Genome Equivalents. Zoo. Biol. 8:111-123.

Lacy RC, Alaks G, Walsh A, 1996. Hierarchical analysis of inbreeding depression in Peromyscus polionotus. Evolution 50: 2187–2200.

MacCluer J, Boyce B, Dyke L, Weitzkamp D, Pfenning A, Parsons C, 1983. Inbreeding and pedigree structure in Standardbred horses. J. Hered. 74: 394-399.

Maignel L, Boichard D, Verrier E, 1996. Genetic variability of French dairy breeds estimated from pedigree information. Interbull Bull 14: 49-54.

Man WYN, Nicholas FW, James JW, 2007. A pedigree-analysis approach to the descriptive epidemiology of autosomal-recessive disorders. Prev. Vet. Med. 78: 262-273.

Meuwissen TI, Luo Z, 1992. Computing inbreeding coefficients in large populations, Genet Sel Evol 24: 305-313.

Nei M, 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, 512 pp.

Pérez-Enciso M, 1995. Use of the uncertain relationship matrix to compute effective population size, J. Anim. Breed. Genet. 112: 333-340.

Quaas RL, 1976. Computing the diagonal elements of a large numerator relationship matrix. Biometrics 32: 949-953.

Robertson A, 1953. A numerical description of breed structure. J Agric Sci 43: 334-336.

Vassallo JM, Díaz C, García-Medina JR, 1986. A note on the population structure of the Avileña breed of cattle in Spain, Livest Prod Sci 15: 285-288.

Woolliams JA, Pong-Wong R, Villanueva B, 2002. Strategic optimisation of short and long term gain and inbreeding in MAS and non-MAS schemes, in: Proc. 7thWorld Cong. Genet. Appl. to Livest. Prod., Montpellier, 19_23 August 2002, INRA, Castanet-Tolosan, France, CD-Rom, comm. No. 23_02.

Wright S, 1931. Evolution in mendelian populations. Genetics 16: 97-159.

Wright S, 1978. Evolution and the genetics of populations: Vol. 4. Variability within and among natural populations. University of Chicago Press: Chicago. USA

5.- Acknowledgements

This document has been developed in a collaborative effort which has involved the Breeders Association of Xalda Sheep of Asturias (http:/www.xalda.com/) within conservation project partially funded by grants from INIA, no. RZ02-020 and RZ2004-00007-C02. The authors would like to thank Luis J. Royo, Isabel Álvarez and especially Iván Fernández, for their kind support and help.

All of the comments and suggestions on ENDOG have been greatly appreciated; the suggestions by Antonio Molina have lead to improvements in the manner in which the program manages the genealogical F-statistics; a productive discussion on the mean and interpretation of the genealogical F-statistics has been carried out with Jordi Jordana and

Josep Marmí. We thank Pascal Schaer from the University of Lausanne for an in-depth discussion on the reliability of the N_e estimations from increase of inbreeding carried out by ENDOG. Rodolfo (Fito) Cantet has insisted on including N_e estimations from family size variances (though much more simple than those proposed). Oscar González Recio has fully contributed to a deep understanding of the practical implications of the individual increase in inbreeding coefficients. Ino Curik, from the University of Zagreb, made extensive reports of the interest of ENDOG at both the user and the scientific levels. Andy Dell, of the Department of Biological Science, at the University of Lincoln UK, has helped with editing the English language version of the ENDOG v4.8 user guide.

6.- Published papers using ENDOG

Valera, M., Molina, A., Gutiérrez, J.P., Gómez, J., Goyache, F. (2005) Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. *Livestock Production Science*, 95: 57-66.

Gutiérrez, J.P., Marmí, J., Goyache, F., Jordana, J. (2005) Pedigree information reveals moderate to high levels of inbreeding and a population genetic structure in the Catalonian donkey breed. *Journal of Animal Breeding and Genetics*, 122: 378–386.

Ruiz-Flores, A., Nuñez-Dominguez, R., Ramirez-Valverde, R., Dominguez-Viveros, J., Mendoza-Dominguez, M., Martinez-Cuevas, E. (2006) Levels and effects of inbreeding on growth and reproductive traits in Tropicarne and Brown Swiss cattle. *Agrociencia,* 40: 289-301.

Alfonso, L., Parada, A., Legarra, A., Ugarte, E., Arana, A. (2006) The effects of selective breeding against scrapie susceptibility on the genetic variability of the Latxa Black-Faced sheep breed. *Genetics Selection Evolution*, 38: 495-511.

Cecchi, F., Ciampolini, R., Ciani, E., Matteoli, B., Mazzanti, E., Tancredi, M., Presciuttini, S. (2006) Demographic genetics of the endangered Amiata donkey breed. *Italian Journal of Animal Science*, 5, 387-391.

Royo, L.J., Álvarez, I., Gutiérrez, J.P., Fernández, I., Goyache, F. (2007) Genetic variability in the endangered Asturcón pony assessed using genealogical and molecular information. *Livestock Science*, 107, 162-169.

González-Recio, O., López de Maturana, E., Gutiérrez, J.P., (2007) Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *Journal of Dairy Science*, 90, 5744-5752.

Olsson, O., (2007) Genetic origin and success of reintroduced white storks. *Conservation Biology*, 21, 1196-1206.

Kim, S.H., Cheng, K.M.-T., Ritland, C., Ritland, K., Silversides, F.G., (2007) Inbreeding in Japanese quail estimated by pedigree and microsatellite analyses. *Journal of Heredity*, 98, 378-381.

Álvarez, I., Royo, L.J., Gutiérrez, J.P., Fernández, I., Arranz, J.J., Goyache, F. (2007) Genetic diversity loss due to selection for scrapie resistance in the rare Spanish Xalda sheep breed. *Livestock Science*, 111, 204–212.

Cervantes, I., Molina, A., Goyache, F., Gutiérrez, J.P., Valera, M. (2008) Population history and genetic variability in the Spanish Arab Horse assessed via pedigree analysis. *Livestock Science*, 113, 24-33.

Álvarez, I., Royo, L. J., Gutiérrez, J. P., Fernández, I., Arranz, J. J., Goyache F. (2008) Relationship between genealogical and microsatellite information characterising losses of genetic variability: empirical evidence from the rare Xalda sheep breed. *Livestock Science* 115, 80-88, doi:10.1016/j.livsci.2007.06.009.

Gutiérrez, J.P., Cervantes, I., Molina, A., Valera, M., Goyache, F., (2007) Individual increase in inbreeding allows estimating realised effective sizes from pedigrees. *Genetics Selection Evolution*, accepted for publication.

Álvarez, I., Royo, L. J., Gutiérrez, J. P., Fernández, I., Arranz, J. J., Goyache F. (2008) Relationship between genealogical and microsatellite information characterising losses of genetic variability: empirical evidence from the rare Xalda sheep breed. *Livestock Science* 115, 80-88, doi:10.1016/j.livsci.2007.06.009.

Galbusera, P.H.A., Gillemot, S. (2008) Polymorphic microsatellite markers for the endangered golden-headed lion tamarin, Leontopithecus chrysomelas (*Callitrichidae*). *Conservation Genetics* 9, 731-733.

Gutiérrez, J.P., Cervantes, I., Molina, A., Valera, M., Goyache, F. (2008) Individual increase in inbreeding allows estimating realised effective sizes from pedigrees. *Genetics Selection Evolution*, 40, 359-378.

Cervantes, I., Goyache, F., Molina, A., Valera, M., Gutiérrez, J.P., (2008) Application of individual increase in inbreeding to estimate effective sizes from real pedigrees. *Journal of Animal Breeding and Genetics*, 125, 301-310.

Ron Garrido, L., Birchmeier, A.N., Munilla, S., Cantet, R.J.C. (2008) Estimation of effective population size using bivariate discrete distributions for modeling family size in beef cattle. *Livestock Science*, 117, 43-51

Malhado, C.H.M., Ramos, A.D.A., Carneiro, P.L.S., Azevedo, D.M.M.R., Martins Filho, R., De Souza, J.C. (2008) Melhoramento e estrutura populacional em bubalinos da raça Mediterrâneo no Brasil. *Pesquisa Agropecuaria Brasileira*, 43, 215-220.

Moore, J.A., Nelson, N.J., Keall, S.N., Daugherty, C.H. (2008) Implications of social dominance and multiple paternity for the genetic diversity of a captive-bred reptile population (tuatara). *Conservation Genetics*, 9, 1243-1251.

Van Coillie, S., Galbusera, P., Roeder, A.D., Schempp, W., Stevens, J.M.G., Leus, K., Reinartz, G., Pereboom, Z. (2008) Molecular paternity determination in captive bonobos and the impact of inbreeding on infant mortality. *Animal Conservation*, 11, 306-312.

Ghafouri-Kesbi, F., Eskandarinasab, M, Hassanabadi, A. (2008) Investigation of genetic variability and inbreeding characteristics in a population of Zandi sheep. *Canadian Journal of Animal Science*, 88, 409-417.

Kucinskiene, J., Subaciute, V., Kvalkauskas, J., Kucinskas, A., Januskevicius, A., Jokubka, R. (2008) Assessment of blood biochemistry, morphology, feeding quality and mating system in Lithuanian hound population. *Veterinarija ir Zootechnika*, 44, 56-61.

Gómez, M.D., Valera, M., Molina, A., Gutiérrez, J.P., Goyache, F. (2009) Assessment of inbreeding depression for body measurements in Spanish Purebred (Andalusian) horses. *Livestock Science*, 122, 149-155.

Rafat, S.A., Allain, D., de Rochambeau, H. (2009) Genetic description of a divergent selection experiment in Angora rabbits with overlapping generations. *Journal of Animal Breeding and Genetics*, 126, 189-197.

Teegen, R., Edel, C., Thaller, G. (2009) Population structure of the Trakehner Horse breed. *Animal*, 3, 6-15.

Malhado, C.H.M., Carneiro, P.L.S., Pereira, D.G., Martins, R. (2009) Genetic progress and population structure in Nellore cattle in Bahia State, Brazil. *Pesquisa Agropecuaria Brasileira*, 43, 1163-1169.

Gutiérrez, J.P., Cervantes, I., Goyache, F. (2009) Improving the estimation of realised effective population sizes in farm animals. *Journal of Animal Breeding and Genetics*, 126, 327-332.

Gómez, M.D., Goyache, F., Molina, A., Valera, M. (2009) Sire x stud interaction for body measurement traits in Spanish Purebred horses. *Journal of Animal Science,* 87, 2502-2509.

Carneiro, P.L.S., Malhado, C.H.M., Martins, R., Carneiro, A.P.S., Silva, F.F.E., Torres, R.D. (2009) The Indubrasil breed in the Brazilian Northeast: breeding and population structure. *Revista Brasileira de Zootecnia-Brazilian Journal of Animal Science*, 38, 2327-2334.

Melucci, L.M., Birchmeier, A.N., Cappa, E.P., Cantet, R.J.C. (2009) Bayesian analysis of selection for greater weaning weight while maintaining birth weight in beef cattle. *Journal of Animal Science,* 87, 3089-3096.

Malhado, C.H.M., Carneiro, P.L.S., Martins, R., Azevedo, D.M.M.R. (2009) Genetic and populational background of Pure Nelore cattle breed in Brazilian Northeastern Sertao. *Pesquisa Agropecuaria Brasileira*, 44: 713-718.

Cervantes, I., Gutiérrez, J.P., Molina, A., Goyache, F., Valera, M. (2009) Genealogical analyses in open populations: the case of three Arab-derived Spanish horse breeds. *Journal of Animal Breeding and Genetics*, 126, 335-347.

Goyache, F., Fernández, I., Espinosa, M.A., Payeras, L., Pérez-Pardal, L., Gutiérrez, J.P., Royo, L.J., Álvarez, I. (2010) Análisis demográfico y genético de la raza ovina Mallorquina. *ITEA*, 106, 3-14.

Nagy, I., Curik, I., Radnai, I., Cervantes, I., Gyovai, P., Baumung, R., Farkas, J., Szendro, Z. (2010) Genetic diversity and population structure of the synthetic Pannon White rabbit revealed by pedigree analyses. *Journal of Animal Science,* 88, 1267-1275.

Bartolomé, E., Goyache, F., Molina, A., Cervantes, I., Valera, M., Gutiérrez, J.P. (2010) Pedigree estimation of the (sub) population contribution to the total gene diversity: the horse coat colour case. *Animal*, 4, 867-875.

Panetto, J.C.C., Gutiérrez, J.P., Ferraz, J.B.S., Cunha, D.G., Golden, B.L. (2010) Assessment of inbreeding depression in a Guzerat dairy herd: effects of individual increase in inbreeding coefficients on production and reproduction. *Journal of Dairy Science*, 93, 4902-4912.

Pedrosa, V.B., Santana, M.L., Oliveira, P.S., Eler, J.P., Ferraz, J.B.S. (2010) Population structure and inbreeding effects on growth traits of Santa Ines sheep in Brazil. *Small Ruminant Research*, 93, 135-139.

Maki, K. (2010) Population structure and genetic diversity of worldwide Nova Scotia Duck Tolling Retriever and Lancashire Heeler dog populations. *Journal of Animal Breeding and Genetics*,127, 318-326.

Flury, C., Tapio, M., Sonstegard, T, Droegemueller, C., Leeb, T, Simianer, H., Hanotte, O., Rieder, S. (2010) Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *Journal of Animal Breeding and Genetics*, 127, 339-347.

Peixoto, M.G.C.D., Poggian, C.F., Verneque, R.S., et al. (2010) Genetic basis and inbreeding in the Brazilian Guzerat (Bos indicus) subpopulation selected for milk production *Livestock Science*, 131, 168-174.

Santana, M.L., Oliveira, P.S., Pedrosa, V.B., Eler, J.P., Groeneveld, E., Ferraz, J.B.S. (2010) Effect of inbreeding on growth and reproductive traits of Nellore cattle in Brazil. *Livestock Science*, 131, 212-217.

Jordana, J., Marmi, J., Avellanet, R., Aranguren-Méndez, J.A., Ferrando, A., Goyache, F. (2010) Molecular, genealogical and morphometric characterisation of the Pallaresa, a Pyrenean relic cattle breed: insights for conservation. *Livestock Science*, 132, 65-72.

Malhado, C.H.M., Carneiro, P.L.S., Malhado, A.C.M., Martins, J.A.M., Martins, R., Bozzi, R. (2010) History of registered Gyr breed in Brazilian Northeast: population structure and genetic improvement of growth traits. *Ciencia Rural*, 40, 1385-1391.

Postma, E., Martini, L., Martini, P. (2010) Inbred women in a small and isolated Swiss village have fewer children. *Journal of Evolutionary Biology*, 23, 1468-1474.

Boettcher, P.J., Tixier-Boichard, M., Toro, M.A., Simianer, H., Eding, H., Gandini, G., Joost, S., Garcia, D., Colli, L., Ajmone-Marsan, P. (2010) Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Animal Genetics*, 41, 64-77. Suppl. 1.

Cervantes, I., Goyache, F., Molina, A., Valera, M., Gutiérrez, J.P. (2010) Estimation of effective population size from the rate of coancestry in pedigreed populations. *Journal of Animal Breeding and Genetics*, in press.