Transformer-3

A program for the analysis of molecular population genetic data

User's Manual

Juli Caujapé-Castells
&
Mario Baccarani-Rosas

EXEGEN SOFTWARE

Transformer - 3

User's manual

by

Juli Caujapé-Castells and Mario Baccarani-Rosas

Laboratorio de Biodiversidad Molecular y Banco de DNA
Jardín Botánico Canario "Viera y Clavijo"
Ap. de correos 14 de Tafira Alta
35017 Las Palmas de Gran Canaria (Spain)

&

EXEGEN SOFTWARE
C. Simancas 32, 1º
35010 Las Palmas de Gran Canaria (Spain)

ON THE Transformer PROJECT AND *Transformer-3*

The Transformer project aims at streamlining the generation, storage, interpretation, processing and application of molecular population genetic data, especially as related to Biological Conservation.

Transformer-3 is one computer program within the Transformer project. It allows the user to concentrate in the accurate interpretation of molecular patterns and in the discussion of quantitative results through automating data transformations and analyses that are otherwise burdensome, complex and prone to error.

Through saving research time while increasing accuracy, *Transformer-3* permits the effective implementation of urgency in the growing number of practical applications of molecular population genetic information.

CREDITS

The Transformer project was conceived and developed by Juli Caujapé-Castells while he was responsible for the molecular biodiversity labs and DNA bank at the Jardín Botánico Canario "Viera y Clavijo" (since 1999 until present) and a "Ramón y Cajal" researcher in this institution (since 2001 until present).

Transformer-3 has been programmed by Mario Baccarani Rosas, and is the result of a collaborative effort between the Jardín Botánico Canario "Viera y Clavijo" and the Department of Software of EXEGEN.

Transformer-3 was developed in the absence of funding.

GENERAL CHARACTERISTICS OF *TRANSFORMER-3*

Transformer-3 is programmed in visual basic using a Microsoft Excel® sheet, so it will run in any PC that can contain the Microsoft Office® package.

This program is suitable for codominant (allozyme or microsatellite) and dominant (AFLP, RAPD, RFLP) data for up to 66,000 diploid individuals distributed in a maximum of 50 populations.

Use the program only in PCs. Since *Transformer-3* runs macros, check that your macro security is set at least to medium, and also make sure that you do not have Excel add-ins in your computer (as these might interfere with some commands in *Transformer-3*).

DISCLAIMER

Transformer-3 can be downloaded from EXEGEN's website (WEBSITE OF EXEGEN) without charge, and may be distributed freely if and when (i) it does not undergo any modification, (ii) this manual and the three example files are attached without changes, and (iii) it is adequately cited in all papers and communications.

Transformer-3 is provided «as is» without any kind of warranty. In no case will the authors or their supporting institutions be liable for any trouble resulting from the use of this software or of its accompanying documentation.

Questions, suggestions, criticisms and bug reports on *Transformer-3* are very much welcome. Address them to

exegensoftware@exegen.org

DO CITE *TRANSFORMER-3* IF YOU USE IT

No one is obliged to download *Transformer-3*. Therefore, if you use this program, please cite it. This is how:

Caujapé-Castells J, Baccarani-Rosas M (2005) *Transformer-3*: a program for the analysis of molecular population genetic data. Exegen software.

The support we receive through the citations is very important in order to facilitate our seeking the necessary funds to improve the program further.

FUTURE RELEASES

Transformer-3 is already being improved to offer a much wider range of possibilities. We hope that *Transformer-4* will include image processing and interpretation routines, together with much enhanced calculation capabilities, while keeping the versatility of the present version. All the persons that download *Transformer-3* will be included in a mailing list, so that they will be notified of any news concerning *Transformer-3* and the future developments of the Transformer project.

ACKNOWLEDGEMENTS

We thank all the biologists and students at the Jardín Botánico Canario "Viera y Clavijo", who were indeed the ones that triggered the creation of the program and the first to provide data and feedback to streamline it, especially Carolina Suárez García, Olga Fernández Palacios, Sara Mora Vicente, Blas Vilches Navarrete, Felicia Oliva Tejera, Magui Olangua Corral, Juan Luis Sánchez Doreste and Moisés Soto.

We are much indebted to Eugenio Reyes (head of the Ethnobotany bank of the JBCVC) for his encouragement, support and generosity.

We also thank the people that have helped us through their continued friendship, support or scientific stimulation, especially Julia Pérez de Paz, Rosa Febles, David Bramwell, Alicia Roca, Bernardo Navarro, Pepe Naranjo, Águedo Marrero, Juan Manuel López Ramírez and Pepa Navarro (at the Jardín Botánico Canario "Viera y Clavijo"), Juan Francisco Rodríguez, Joaquín Ocón, Yizzat Sabbagh, Gonzalo Piernavieja and Julieta Schallenberg (at the Instituto Tecnológico de Canarias), Miguel González-Pérez and Pedro Sosa (at the Universidad de Las Palmas de Gran Canaria), Pilar Catalán (at the Universidad de Zaragoza), Jerzy T. Puchalski (at the Polish Academy of Sciences) and Juan Mota (at the Universidad de Almería).

Finally, we thank all the scientists outside the JBCVC that have tested the beta versions of *Transformer-3* and have provided suggestions and/or feedback. These include César Blanché, María Bosch, Joan Simón and Jordi López Pujol (at the Departament de Botànica de la Facultat de Farmàcia of the Universitat de Barcelona, Spain), Andrea Blanquer and Yosune Uriz (at the Centre d'Estudis Avançats de Blanes, Spain), Molly Stephens (at the Genomic Variation Lab, UC Davis, USA), Christopher Richards (at the National Center for Genetic Resources Preservation, Colorado State University, USA), Alisa P. Ramakrishnan (at the Department of Biology of Portland State University, USA), Javier González (at the Institut für Pharmazie und Molekulare Biotechnologie of the Ruprecht-Karls-Universität Heidelberg, Germany), Dorothée Baïlle (at the Laboratoire d'Evolution du Développement des Nématodes of the Institut Jacques Monod, France), Mikael Lönn (at the Södertörn

University College, Sweden), Anne Chenuil-Maurel (at the CRNS Université de Marseille and Station Marine d'Endoume, France), Colin Kelleher (at the Genome BC - Forestry Genomics department of the University of British Columbia, Canada), Vahid Erfani Moghadddam (at the Biotechnology Lab of the Gorgan University of Agricultural Science, Iran), Rik van Wijk (at the Wageningen University, The Netherlands), John Gaskin (at the USDA-ARS-NPARL in Sidney, Australia), Melinda Peters (at the Department of Botany of North Carolina State University, USA), Karen Mock (at the Forest, Range, and Wildlife Sciences Dept. of Utah State University, USA), Manoj P. Brahmane, (at the Central Inland Fisheries Research Institute of Barrackpore, India), James M. Bradeen (at the Department of Plant Pathology of the University of Minnesota, USA), Cherdsak Liewlaksaneeyanawin (at the Dept. of Forest Sciences of the U. of British Columbia, Canada), Jennifer Ramp (at the University of Colorado Museum, USA), Alexander Krings (at the Department of Botany of North Carolina State University, USA), and Peter Galbusera (at the Center for Research and Conservation of the Royal Zoological Society of Antwerp, Belgium).

PREFACE

The history behind *Transformer-3* is, in short, another one of chance and necessity. Necessity came along with the growing bulk of data analyses related to the population genetic projects under way at the Jardín Botánico Canario «Viera y Clavijo» (JBCVC), that triggered the creation of a *Transformer-1* (Caujapé-Castells 2001). That first version proved to be suitable enough to bypass a number of burdensome and error-prone aspects of molecular population genetic data analysis, though it was still too tangled to be released without shame. Therefore, *Transformer-1* was only operated by Juli in his personal computer. Available time was in very short supply since the completion of *Transformer-1*, and this alone would have provided Juli an excellent excuse not to pursue a better version; after all, that first program already analysed the molecular population genetic data generated at the JBCVC much faster than usual.

Perhaps the chance to develop a *Transformer-2* wouldn't have turned up had it not been by Eugenio Reyes, an educator at the JBCVC, who was aware of the Transformer project. This friend put Juli in contact with the researchers at the División de Software of the Instituto Tecnológico de Canarias (ITC), who were very receptive to the idea of helping him develop a better Transformer that implemented fully the project he had in mind. After several meetings and unsuccessful trials, the challenge was undertaken by Mario, who programmed *Transformer-2* and made possible many ideas that were just starving for opportunity.

Transformer-3 has also been programmed by Mario and was produced by the Department of Software of Exegen in the absence of funding. This new version incorporates enhanced transformation capabilities for co-dominant data (it deals with seven new programs of analysis), but it works with dominant data matrices as well (AFLPs, RAPDs or RFLPs). Thus, *Transformer-3* is a universal software for molecular population genetic data analysis.

Like its predecessors, *Transformer-3* saves a lot of time, thereby allowing the researcher to concentrate on what really matters: the critical discussion of results and hypotheses.

EXEGEN software
Las Palmas de Gran Canaria, October 13th 2005

INDEX

Section 1. Entering Data
- 1.1. Generalities on the drawing utility
  - 1.1.1. Advantages of drawing the interpretations
  - 1.1.2. General features of the drawing matrix
  - 1.1.3. Drawing heterozygous individuals
  - 1.1.4. Adding sheets to your drawing project
  - 1.1.5. Phantom bands and missing data
- 1.2. Drawing your interpretations
  - 1.2.1. Getting started
  - 1.2.2. Choosing population codes
  - 1.2.3. Define the enzyme/primer
  - 1.2.4. Place the alleles in their correct positions
  - 1.2.5. Draw the alleles
  - 1.2.6. Inserting new alleles in the drawing matrix
  - 1.2.7. Inserting new loci
  - 1.2.8. Drawing "phantom" bands
  - 1.2.9. Inserting new individuals
  - 1.2.10. Get the genotypes from the drawing
  - 1.2.11. Tips on drawing the data
- 1.3. Entering a table of genotypes
  - 1.3.1. How to input your genotypes for analysis

Section 2. Configuring the data
- 2.1. The default configuration
- 2.2. Dealing with groups of populations
  - 2.2.1. Analysing population subsets
  - 2.2.2. Comparing independent groups of populations
  - 2.2.3. Including populations in more than one group

Section 3. Processing the data
- 3.1. The Biosys format
  - 3.1.1 Obtaining the Biosys format
- 3.2. The Bottleneck format
  - 3.2.1. Obtaining the Bottleneck format
- 3.3. The GenePop format
  - 3.3.1. Obtaining the Genepop format
- 3.4. The GeneStat format
  - 3.4.1. Obtaining the GeneStat format
- 3.5. The Ntsys format
  - 3.5.1. Ntsys format requirements
  - 3.5.2. Obtaining the Ntsys format
- 3.6. The PopGene format
  - 3.6.1. Obtaining the PopGene format
- 3.7. The AFLPsurv format
  - 3.7.1. Obtaining the AFLPsurv format
- 3.8. The Arlequin format
  - 3.8.1. Obtaining the Arlequin format
- 3.9. The FSTAT format
  - 3.9.1. Obtaining the FSTAT format
- 3.10. The GDA format
  - 3.10.1. Obtaining the GDA format

3.11. The GenAlex format
3.11.1. Obtaining the GenAlex format
3.12. The PASSAGE format
3.12.1. Obtaining the PASSAGE format
3.13. The SPAGeDI format
3.13.1. Obtaining the SPAGeDI format
3.14. The probabilities of allelic loss
3.14.1. On the probability of loss
3.14.2. Obtaining the probabilities of allelic loss
3.15. Tips on processing the data
LITERATURE CITED
APPENDIX: The example files.

Section 1. Entering data

Transformer-3 offers a versatile, interactive data entry interface that makes corrections and manipulations easy to implement. At present, you can feed *Transformer-3* with the drawings of the interpretations of your molecular patterns (see section 1.1) or with a matrix of genotypes that you have to type (see section 1.2).

1.1 GENERALITIES ON THE DRAWING UTILITY

Transformer-3 allows you to store your interpretations in an interactive drawing matrix. This tool allows the user

1. To have a permanent record of the interpretations of molecular patterns that can be easily modified and corrected.

2. To generate a genotype file for any combination of loci, which is the basis for any subsequent data transformations and analyses.

Although we believe that drawing interpretations is advisable in most cases (particularly if using allozymes), it is especially so if you begin to interpret your molecular patterns right when you obtain the first consistent data.

Building your database little by little is practically effortless and allows you to track eventual changes and check previous interpretations easily while saving a lot of time and errors.

1.1.1 Advantages of drawing the interpretations

The major advantage of drawing the interpretations is that, once you are done, quantitative data for any possible configuration of populations and loci will be a few easy clicks away (see sections 2 and 3).

However, there are at least three more powerful reasons to use this tool of *Transformer-3*.

1. You can **forget about genotyping individuals**, as *Transformer-3* will do it for you (see section 1.2.10). Therefore, you are less prone to make the mistakes that are so frequent when you interpret by hand.

2. You may **correct or modify your interpretations** (see sections 1.2.6 to 1.2.9) by moving, inserting or deleting any number of individuals, alleles, loci or spaces easily at any point of the interpretation process.

3. You will **have a visual record of the interpretations** which is much easier to scan than a whole table of genotypes and which can be used nicely in presentations (see the attached file «transf-draw.xls»).

1.1.2. General features of the drawing matrix

1. The drawing matrix of *Transformer-3* is conceived to draw the interpretations so that the fastest alleles appear at the left-hand side of the drawing and the slower ones at the right-hand side. To put it formally, the drawing corresponds to the original gel shuffled back to front and then turned 90º counter-clockwise (see Figure 1 for an illustration).

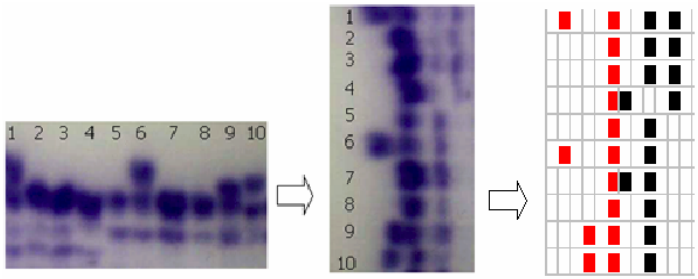


Figure 1. Original picture of an allozyme gel for a monomeric enzyme (left) and how should it appear in the drawing matrix of *Transformer-3* (right). Different colours stand for different loci (red is locus 1 and black is locus 2)

Although this way of drawing may appear counter-intuitive at first, it does not take long to become familiar with it. Its advantages are

that it allows the program to have faster analytical algorithms, while the user can «read» the alleles from left to right in several loci for many individuals.

2. For each enzyme/primer, the drawing utility of *Transformer-3* consists of (see Figure 2)

(a) **an enzyme/primer header** that contains

(i) the name of the enzyme/primer (framed), and

(ii) the positions of the alleles detected

(b) **a drawing matrix**, where you can insert and delete columns to make it fit your molecular patterns

(c) **a genotype area** with the label "Gntp", containing as many columns as loci you have defined for that enzyme/primer (the limit is 10 loci per enzyme/primer).

These columns will be coded according to the enzyme/primer name and will remain empty until you decide to genotype that locus (see section 1.2.10)

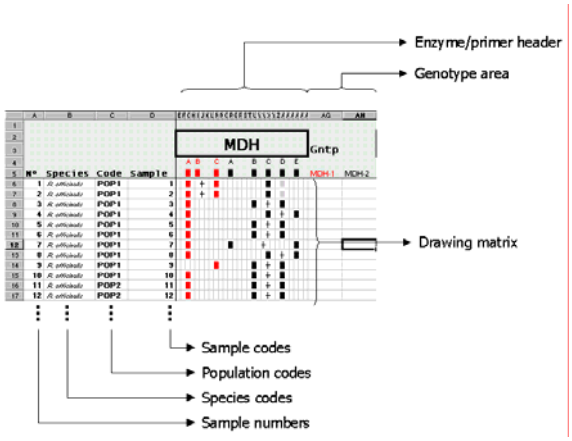
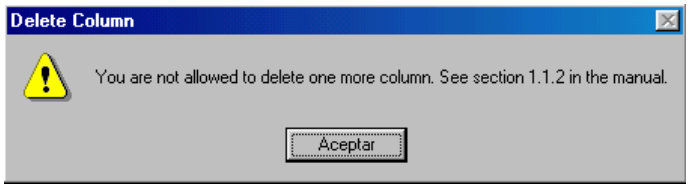


Figure 2. Detail of the drawing matrix of *Transformer-3*. The first four columns of the area coloured in light grey correspond to the sample numbers (N), the species names (Species), the population codes (Code) and the sample codes (Sample). The next columns in grey show the enzyme/primer header for the enzyme MDH (which, in this case, has two loci with two and five alleles, respectively) and the genotype area. The white area below the enzyme/primer header is the drawing matrix, where the user can draw the interpretations of gels following the indications in the manual. Only a part of the drawing matrix for MDH is shown.

3. The drawing matrix for a given enzyme must have at least 11 columns, so that, if you reach this minimum width, *Transformer-3* will not allow you to delete columns (see section 1.2.4), and an error message like the one below will appear



4. For the sake of uniformity, *Transformer-3* assigns a predefined colour to all the alleles belonging to a given locus.

All the alleles of the first locus within an enzyme/primer will be red, those at the second black, those at the third blue, and so on until the tenth, whose alleles are violet (see Figure 3 for the colour codes associated with each locus).

5. *Transformer-3* will only interpret the alleles in the drawing matrix whose colour and position are defined at the enzyme/locus header (see section 1.2.10). The palette of pre-assigned allele colours for the maximum of ten loci is illustrated in Figure 3b.

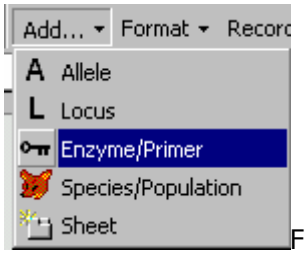


Fig.3a

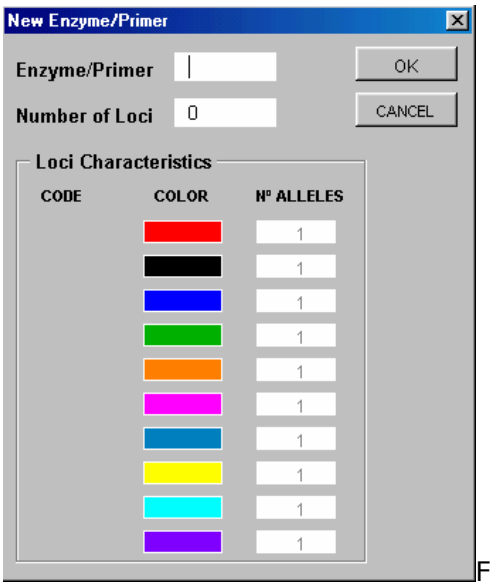


Fig.3b

1.1.3. Drawing heterozygous individuals

Heterozygous individuals in monomeric allozyme loci and in microsatellite loci should be represented by two bands of the same colour (see Figure 4)

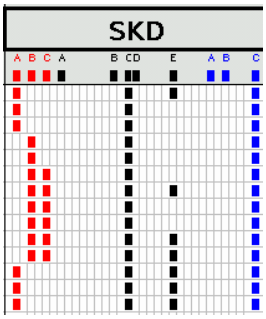


Figure 4. Example showing several heterozygous individuals in the monomeric enzyme SKD (from the "transf-example.xls" file). Note that only the red locus (SKD-1) and the black locus (SKD-2) have heterozygous individuals for this section of the data.

Heterozygous individuals in dimeric and multimeric allozyme loci should be represented by three symbols: the two bands at the extremes should be assigned the corresponding locus colour, and the heterodimer(s) should be a pre-defined symbol (see Figure 5).

After selecting the cell where you want to insert the heterodimer, its symbol can be drawn in one of two ways:

(a) pressing simultaneously "Alt" and "Z", or
(b) pressing the button "heterodimer" in the bar chart menu .

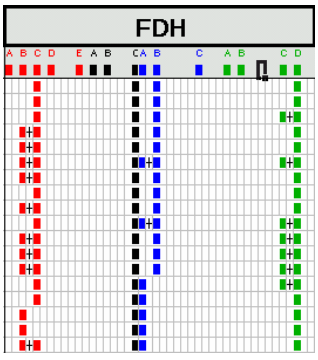


Figure 5. Example of heterozygous individuals in three of the four loci defined for the dimeric enzyme FDH in the "transf-draw.xls" example file (the black locus FDH-2) is monomorphic in this section of the file.

1.1.4. Adding sheets to your drawing

If you have many polymorphic enzyme/primers in your project, it is probable that their interpretations do not fit in a single Excel sheet (Excel has a column number limit which is very short for most data sets). In this case, you can add new sheets selecting the option "sheet" in the button "Add" in the toolbar menu (see Figure 6).

The first sheet will be named sheet 1, the next one sheet 2, and so on up to (eventually) sheet 10.

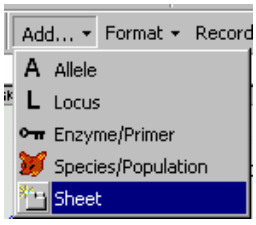


Figure 6. Selecting "Add sheet" from the toolbar menu

The contents of the newly added sheet will be exactly the same as that of the first one, including the drawings. You have to format the new sheet so that it only keeps the species and population codes for your samples. To do this, select the option "Current sheet" in the button "Format" in the toolbar menu (see Figure 7).

Figure 7. Formatting a newly created sheet for drawing new enzyme/primer interpretations.

Be careful not to select "All sheets" within "Format" unless you want to erase the whole contents of your interpretation file.

1.1.5. Phantom bands and missing data

Every locus colour in *Transformer-3* has two associated degraded tones (both of them fainter than the corresponding allele colour) that can be used to draw bands that you do not want to include in the final interpretations (Figure 8) [see section 1.2.8. for details].

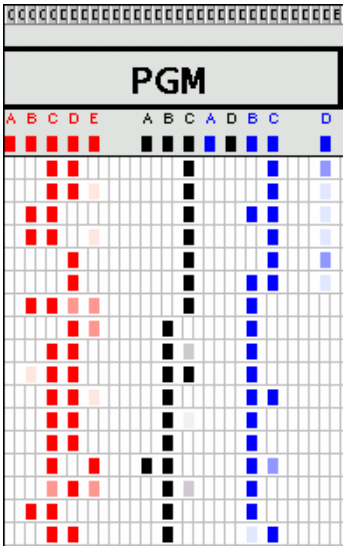


Figure 8. Example of "phantom bands" in a gel corresponding to the enzyme PGM. There are phantom bands in the three loci defined for this enzyme (see section 1.2.8. for details).

Also, if you cannot interpret a given individual for a given locus, you can leave it blank. *Transformer-3* will just add a 999 to the corresponding genotype when it interprets the pattern (the file "transf-draw.xls" contains many individuals with empty loci).

1.2. DRAWING YOUR INTERPRETATIONS

1.2.1. Getting started

a) Select the option "Species/population" of the button "Add" in the toolbar menu (Figure 9a).

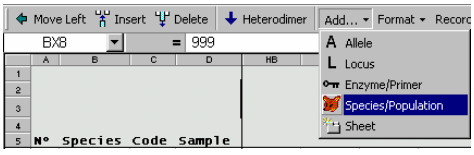


Figure 9a.

b) Introduce the name of the first species you want to include in the drawing file you are about to create, the population code and the number of individuals in that first population in the dialog that will appear (Figure 9b).

Figure 9b. Introducing a population in the *Transformer-3* drawing sheet.

Just feed in this box what you have at present, and do not worry if you plan to include more populations in your project or sample more individuals for a given population; you will be able to add these at any moment of the interpretation process (see section 1.2.6., 1.2.7., and 1.2.9.).

After filling in this box, *Transformer-3* will write automatically in the drawing sheet the number of individuals that you have assigned to each population using four columns (see Figure 10):

1. The first column is the total number of samples
2. The second column is the name of the species you have input
3. The third column is the population code of your choice
4. The fourth column is a numerical free code that you may want to assign in order to identify each individual.

Write only in the fourth column to introduce the individual codes. It is better not to write anything in the other columns.

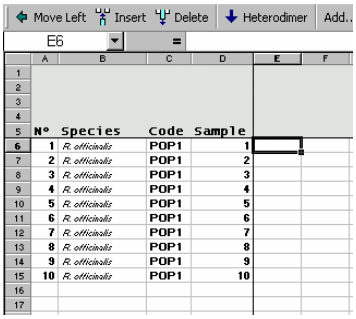


Figure 10. Detail of the drawing sheet after pressing "OK" with the selection made in Fig. 9b.

c) Insert the remaining populations of your project in the *Transformer-3* drawing sheet by selecting "Species" in the "Add" button from the toolbar as many times as needed (see Figure 11a, b and c).

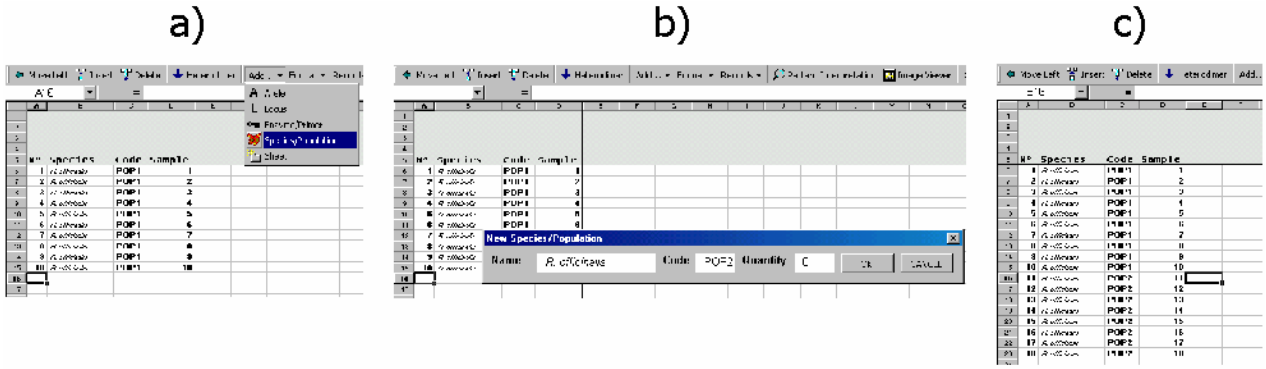


Figure 11. Adding new populations to the *Transformer-3* drawing sheet.

1.2.2. Choosing population codes

For a population code you can use any string of characters, including numbers and signs. The only restriction is to choose codes without empty spaces whose symbols (if any) do not conflict with the entry formats of any of the programs that *Transformer-3* generates files for (see section 2). Some examples of two population codes that *Transformer-3* can deal with are (HILL, LAKE), (HILL1, HILL2), (HILLA, HILLB), (LAKE-SP1, LAKE-SP2), (101-A, 101-B). Have a look at the attached file "transf-example.xls" for other examples.

1.2.3. Define the enzyme/primer

After inserting the samples of your project, you have to define the basic traits of your enzymes/primers.

1) Select «enzyme/primer» in the button «Add» from the toolbar (Figure 12a).

   Then, you will be presented with a menu that asks you to input the basic characteristics of the molecular patterns you're

about to introduce (Figure 12b) in order to configure the loci in the *Transformer-3* drawing sheet.

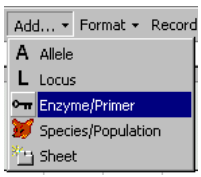


Figure 12a. Adding an Enzyme/Primer to the drawing sheet.

2) Fill in the dialog

If you are starting the molecular interpretations from scratch, just feed the number of alleles you detected in your first gel. Again, do not worry about new alleles, individuals or loci that you may have to add in the future; you will be able to do it easily at any point of the interpretation process (see sections 1.2.6., 1.2.7. and 1.2.9.). In the example in Figure 12b, the enzyme MDH has two loci with 3 and 5 alleles, respectively.

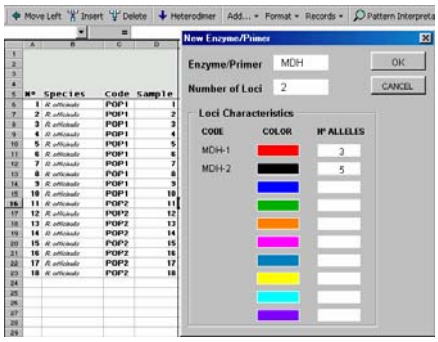


Figure 12b. Defining the basic features of the new Enzyme/primer. In this case (allozymes), the enzyme (MDH) has two loci (MDH-1 and MDH-2) with 3 and 5 alleles, respectively.

If you are drawing allozymes, write

a) the code of the enzyme,

b) the number of loci, and

c) the number of alleles for each locus.

If you are drawing microsatellite profiles, just

a) introduce the primer code in the corresponding cell,

b) put a «1» in the box «number of loci», and

c) introduce the number of alleles you're about to draw.

3) press "OK".

Transformer-3 will then ask you if everything is correct. If you confirm, the number of alleles that you have selected for each locus in a given enzyme/primer will appear automatically below the enzyme/primer header, with their corresponding colour and letter codes. Figure 13 illustrates the default conformation of the drawing matrix for the selection made in Fig. 8. Since we selected 2 loci with 3 and 5 alleles (respectively), there will be 3 red bands and 5 black ones.

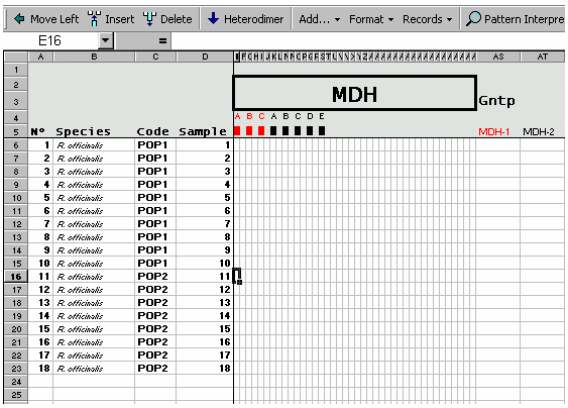


Figure 13. The default conformation of the drawing matrix for the selection made in Fig. 8.

Also, at the right of the drawing matrix for each enzyme/ primer there will be as many columns as loci you have defined, each of them correspondingly coded and coloured. These columns (two in the example) will remain empty until you decide to interpret your patterns (see section 1.2.10.).

1.2.4. Place the alleles in their correct positions

As you can notice in Figure 13, the separation among alleles and loci is assigned automatically by *Transformer-3*, and it will probably not correspond to their real separation on the gels. Thus, a first thing you want to do is to adapt the relative positions of the alleles to reflect their positions in the gel. Do it one allele at a time (starting with the one at the far right) as follows:

a) Select the corresponding coloured cell in the enzyme/primer header

b) Click on «move left» or «move right» in the bar chart menu until you have placed all the alleles in the desired positions

Figure 14 illustrates the end of this process for the default pattern in Figure 13.

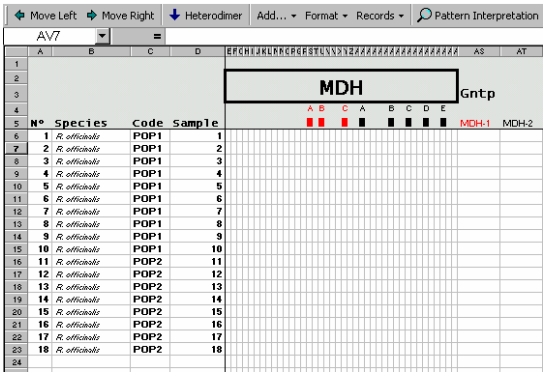


Figure 14. Modification of the pattern in Fig. 13 obtained by moving the alleles to the right. The allele movement began with the black allele labelled "E", followed with the one labelled "D", and so on until the red allele labelled "A"

To insert or delete columns within the drawing matrix,

a) place the pointer at the chosen place in the matrix and

b) press the button «Insert column» or "delete columns" (as needed) in the bar chart menu (Figure 15).

Figure 15

Adding columns is adequate if you need a bigger matrix for drawing the interpretations of your molecular patterns (new columns will be created at the right of the selected cell).

Deleting columns is an option you may want to take in order not to assign more space than strictly needed to represent your interpretations of a given Enzyme/primer (see Figure 16). However, the minimum number of columns in an enzyme/primer is 11. Once you reach this limit, *Transformer-3* will not allow you to delete more columns (see section 1.1.2.3)

You can add or delete columns at any point of the interpretation process. Be careful not to delete a column where you defined an allele. Just in case, *Tranformer-2* will always ask you to confirm the deletion before proceeding.

Figure 16 shows the effect of eliminating the spare columns at the left of the first allele of the red locus in Figure 15.

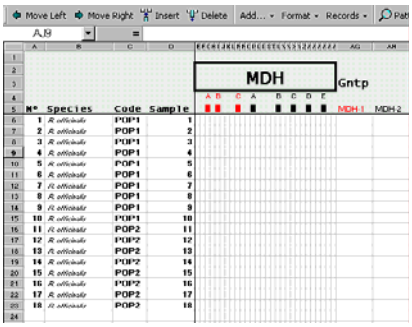


Figure 16. Modification of the pattern in Fig. 14 obtained by deleting columns at the left of the red "A" allele.

For a better visualisation of the patterns, it is advisable to leave at least one blank column between consecutive alleles (see Figure 10). However, *Transformer-3* does not have any problem with interpreting contiguous alleles not separated by a blank column (see Figure 17).

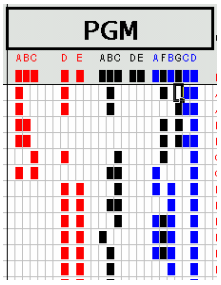


Figure 17. Example of allozyme loci with several contiguous alleles for the enzyme PGM.

1.2.5. Draw the alleles

There is just one possibility to draw an allele for a given individual:

(1) select the cell where you want that allele and press "Alt" and "X" simultaneously

DO NOT draw alleles by copying the coloured cell from the enzyme/primer header and pasting it in the corresponding individual. *Transformer-3* will let you do it, but this can give rise to errors in the subsequent interpretations. Just use the described combination of keys.

If you use *Transformer-3* for drawing microsatellite profiles, take into account that, at present, the program does not take "size of the allele" or "number of motive repetitions" as a variable, so it will just assign an "A" to the smaller allele, a "B" to the second smaller, and so on.

1.2.6. Inserting new alleles in the drawing matrix

Whenever you detect a new allele in a locus, you have to define it first in the enzyme/primer header, or *Transformer-3* will not recognise it as an allele (see the sections 1.2.3 and 1.2.10).

To define the position of a new allele in one of the already existing loci, follow these steps:

a) Select the cell where you want to place the new allele

You can choose any position in the space assigned to alleles in the enzyme/primer it belongs (see Fig. 18a)

b) In the box that will appear, select the locus colour where that allele should be assigned

Transformer-3 asks you this because there are no restrictions on the relative position of any allele within the drawing matrix for a given locus. This means that, for instance, an allele of the first locus (red) can be placed in the middle of two alleles from the third locus (blue) as well as in the middle of two pre-existing alleles for the first locus. Whatever the case, *Transformer-3* will automatically recode the old alleles according to their new relative positions (see Figures 18a to f for examples).

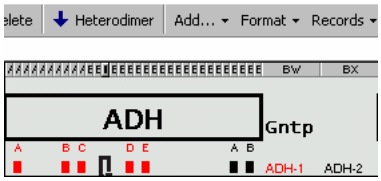


Figure 18a. Select the cell where you want to add a new allele.

Figure 18b. From the custom toolbar, select "Add..." and then "Allele".

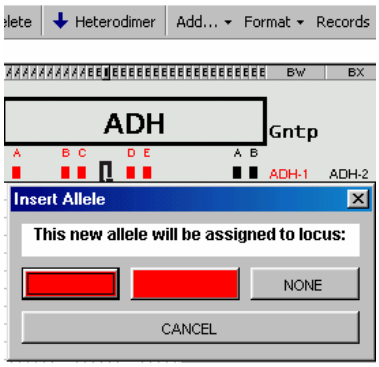


Figure 18c. In the dialog box that appears, the new allele can be assigned either to any of the two flanking loci (only red in the example) or to any other existing locus by selecting [NONE].

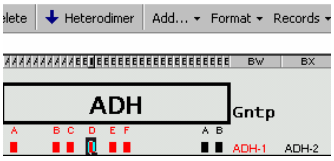


Figure 18d. If we pressed the red button in Fig. 18c, a new allele would be assigned to the red locus (ADH-1) in the selected position.

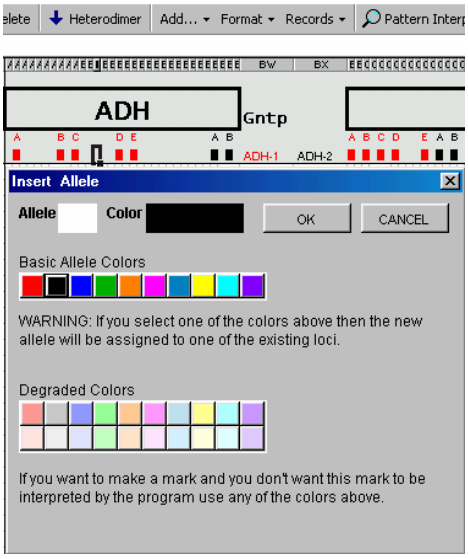


Figure 18 e. If we pressed "NONE" in Figure 18c, a dialog like this would appear. Although all possible allele colours are shown, only selecting the black cell under "Basic allele colours" would insert an allele, because the chosen enzyme (ADH) only has two loci defined. If we choose a colour other than black, an error message would appear.

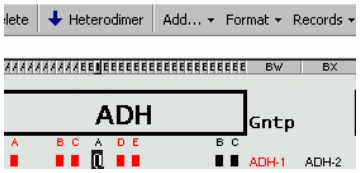


Figure 18 f. After pressing "OK" in 18 e, a new allele appears at the black locus, and the pre-existing alleles at that locus change their codes according to their new position.

Remember that *Transformer-3* only understands diploid data, so that a maximum of two different bands with the locus colour can be used for genotyping an individual at that locus. Just in case, if you draw more than two alleles per locus in a given individual, *Transformer-3* will pop out an error message when it interprets the patterns (see section 1.2.10).

1.2.7. Inserting new loci

If you want to assign a new allele to a new locus that you had not detected in the previous analyses, then just

a) Select the position where you want to place the new locus in the enzyme/primer header (like in Fig. 19a).

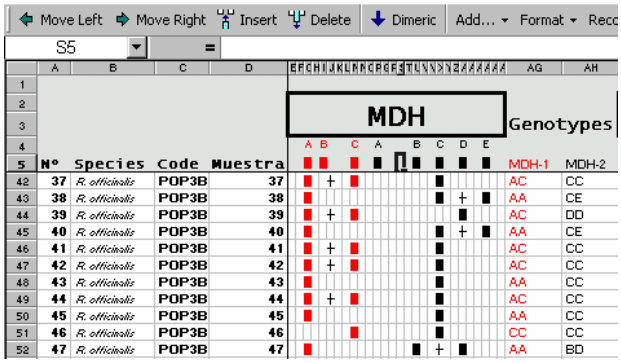


Figure 19a. Select the position where you want to define a new locus.

b) Choose «Add» and then «Locus» in the toolbar menu (see Figures 19b and 19c).

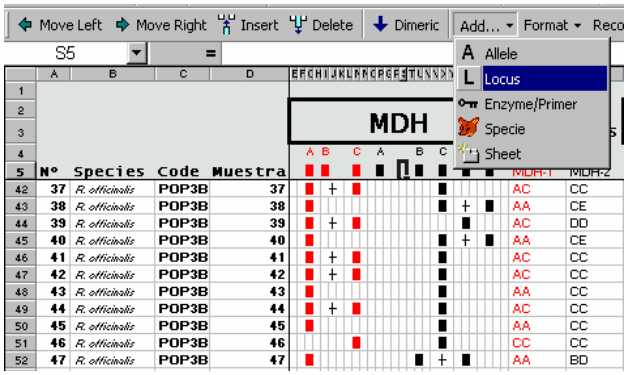


Figure 19 b. Select "Add" and then "locus" from the toolbar menu.

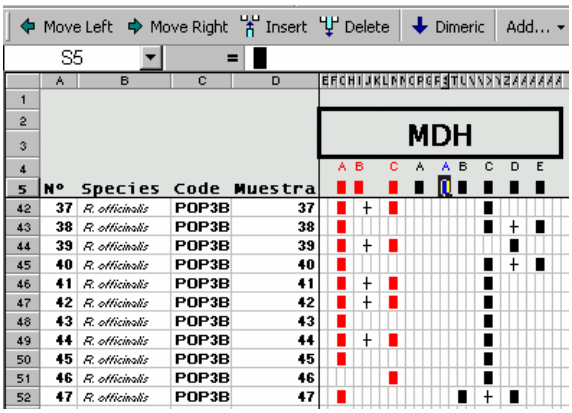


Figure 19c. Transformer automatically inserts the new locus with its corresponding colour.

The new locus can be inserted at any position in the drawing matrix corresponding to a given enzyme/primer. If you place it in the middle of two pre-existing loci, then *Transformer-3* will automatically refurbish the colour codes of the loci at the right of the newly inserted one so that they fit the new conformation. The program will also insert a new column in the genotypes area.

Figure 20 illustrates the effects of the insertion of a new locus between two pre-existing ones (labelled in red and black). *Transformer-3* will re-draw the alleles in black so that they now will belong to the new locus 3 (which should be blue according to the colour code), while those for locus 1 will remain untouched, because they are at the lefthand side of the locus and they are not affected by the appearance of the new locus.

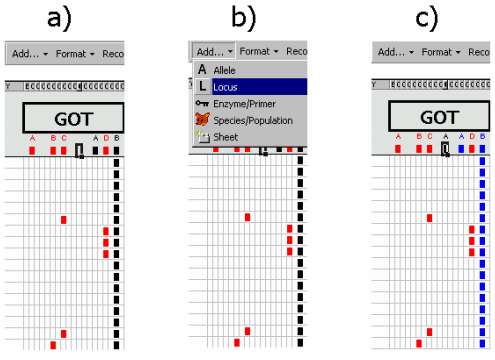


Figure 20. Insertion of a new locus between two pre-existing ones. Note that, after inserting the new locus, if there were any loci at the right side of it, the colours of all their alleles change automatically according to the colour codes in Fig. 3b. In the case of this figure, the alleles in the locus that was previously black changed to blue after the insertion.

1.2.8. Drawing «phantom» bands

Those working with allozymes are used to come across bands whose interpretation is thorny because they cannot be assigned safely to any locus. In most cases, it is convenient to store these so-called «phantom» bands (Arús and Shields 1983) as qualitative information for eventual consideration in the future. At the JBCVC, students that use molecular population techniques are always requested to draw the phantom bands (if any) to purport a more realistic version of the gel that can set the stage to alternative interpretations. This utility can also be used to represent heterodimers.

Transformer-3 offers two degraded versions of the colour codes assigned to each locus to represent these phantom bands (see example in Figure 21).

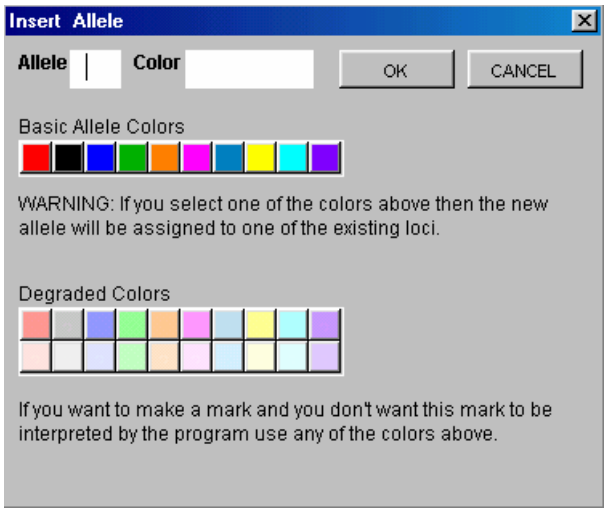


Figure 21. Palette of possible colours for phantom bands (below the heading "Degraded colors") for each of the 10 corresponding loci colours. Notice that each locus colour has two different associated tones that can be used to represent phantom bands. See sections 1.1.5 and 1.2.8 for details on the use of these bands.

1.2.8.1. If you want to insert a phantom band in a position which is defined as an allele in the enzyme/primer header, then

a) Select the position where you want to insert it and
b) Press "ALT" and "C" simultaneously.

A degraded version of the corresponding locus colour will appear in the selected cell. If you press "ALT+C" again, then an even fainter version of the locus colour will appear. If you press "ALT+C" a third time, then the first degradation of the allele colour will appear, and so on.

1.2.8.2. If the phantom band is not in a position defined as an allele in the enzyme/primer header, then (see Figures 22a to g)

a) Select the position where you want to place the phantom band in the enzyme/primer header (Fig. 22a)
b) Press "Add a" and then "Allele" in the toolbar menu (Fig. 22b)
c) Press "none" in the dialog box that will appear (Fig 22c)
d) Select the corresponding degraded tone you want to assign to the phantom band (Fig. 22d and e)
e) Draw the phantom bands (Fig. 22f and g)

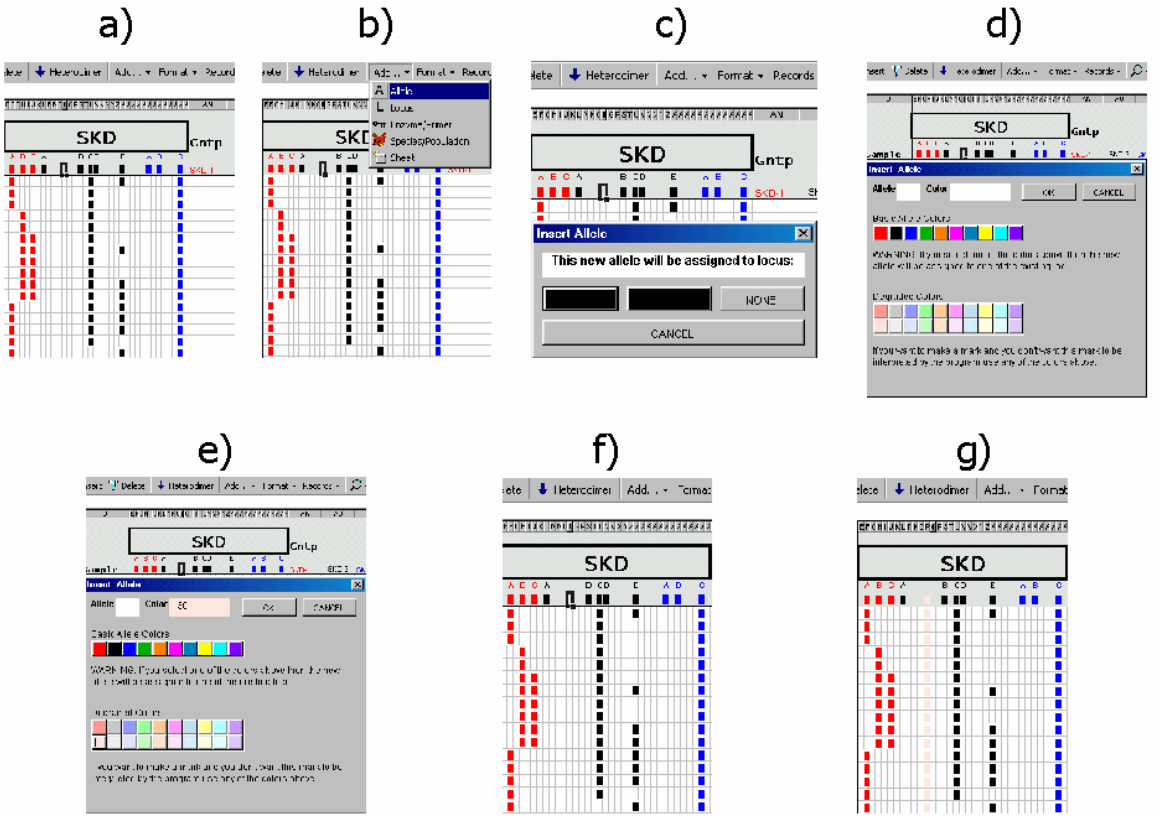


Figure 22. Drawing phantom bands in a position not defined at the enzyme/primer header. In the example, although SKD is an allozyme with uncertain quaternary structure, it is represented as a dimer.

You can draw any number of these «phantom» bands, as *Transformer-3* will not interpret them (remember it only interprets the alleles whose colours and positions coincide exactly with those defined as alleles at the enzyme/primer header).

1.2.9. Inserting new individuals

You can add new individuals to the drawing matrix at any point of the interpretation process. The only small restriction is that, if the newcomers belong to a population that already has representatives in the drawing, they must be added after the last individual for that population.

If a new population is to be added in your project, then *Transformer-3* will do it at the end of the existing file (select "Species/population" in the button "Add" on the toolbar menu and carry on as described in section 1.2.1).

To insert individuals in an already existing population,

a) Select any individual in the population where you want to add the new samples (see Figure 23a)

b) Click on the option «Add» in the button «Records» on the toolbar menu (see Figure 23b)

c) Just in case, *Transformer-3* will remind you you're just about to add new individuals in that population (see Figure 23c)

d) Click «OK» and the pointer will move to the position where it will insert the first of the new individuals (i.e., right after the last one of the pre-existing ones, see Fig. 23d)

e) Write in the dialog box the number of individuals you want to add (Figure 23d).

After completing this process, the corresponding number of cells will appear after the last of the pre-existing individuals in the selected population (see Figure 23e). Notice that, for the newly added individuals (3 in the example), the cells in the column "Sample" are blank, so that you can insert the (eventual) code of the new

samples (their species and population codes will remain the same as for the other individuals from that population). Also notice that the sample codes for the pre-existing individuals will remain the same, but their sample number (N) will have varied according to the number of inserted samples (see Figure 23e and note the changes in N in POP2).

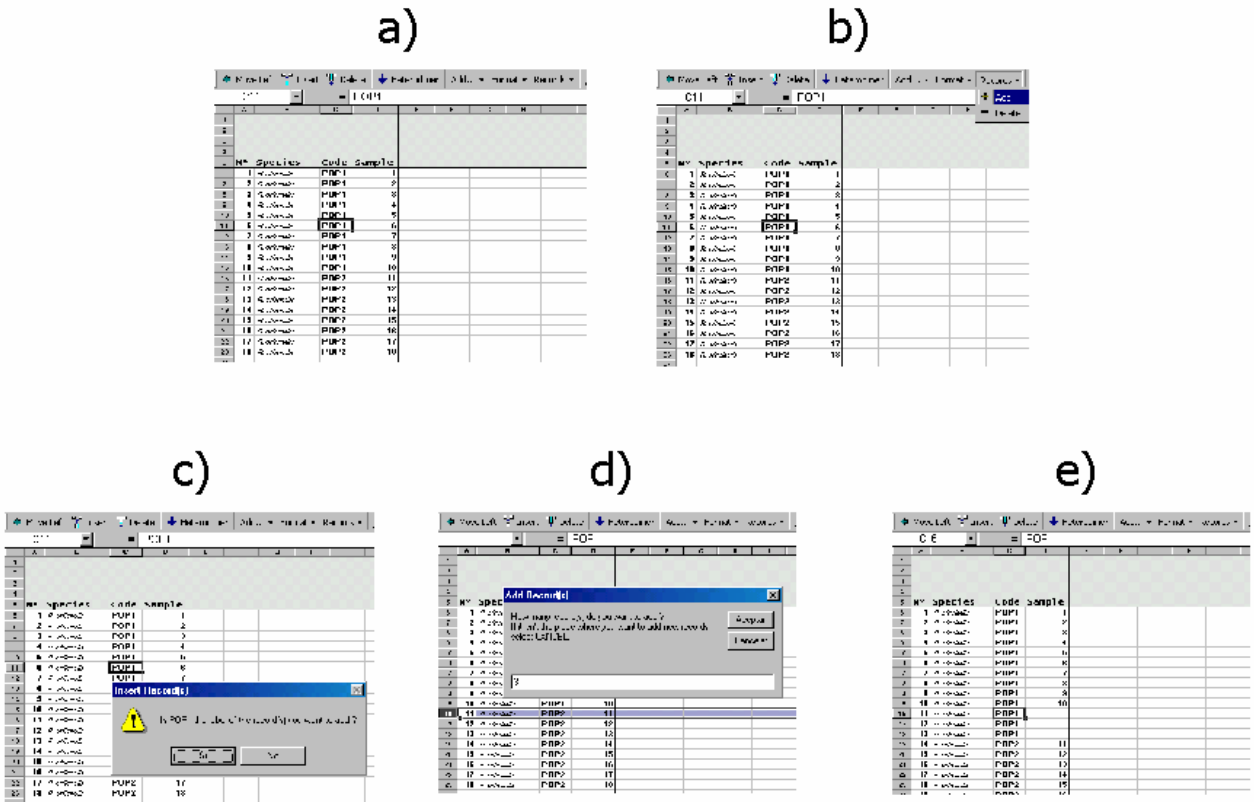


Figure 23. Inserting new individuals in a pre-existing population.

1.2.10. Get the genotypes from the drawing

When you complete the drawings for all the loci and individuals you have included in your project, you are ready to obtain the genotypes.

You first have to request *Transformer-3* to analyse the enzyme/primer patterns one by one. This process allows you to check any possible error more easily than if all the patterns were analysed at once. To analyse the patterns, just

a) Select "Pattern interpretation" in the toolbar menu (Figure 24).

Figure 24

b) Fill in the box that will appear with the name of the enzyme/primer whose pattern you want to analyse (see Figure 25) [the Enzyme/primer to interpret must be in the active sheet].

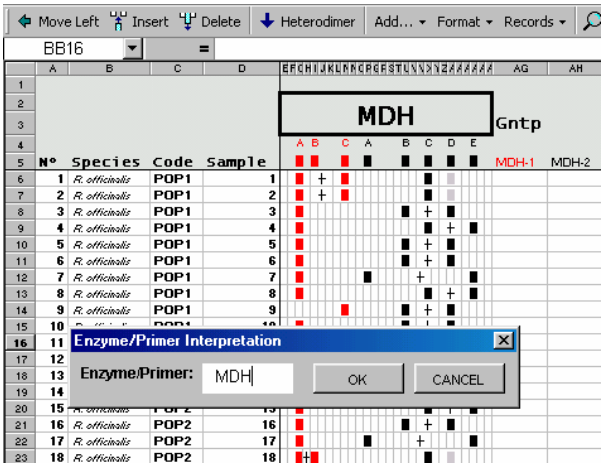


Figure 25. Selecting an enzyme/primer for interpretation

c) Press "OK" and *Transformer-3* will genotype that enzyme/primer (Figure 26).

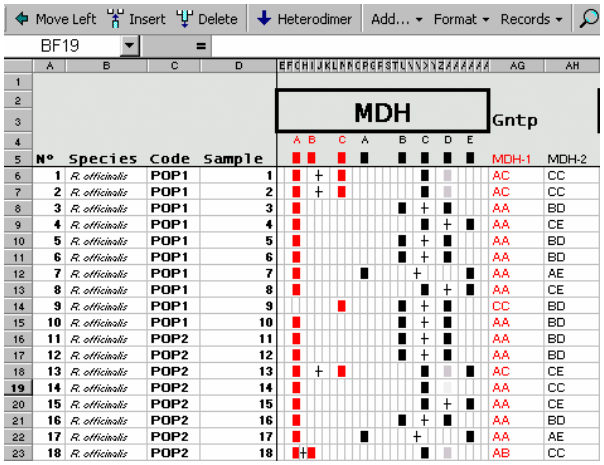
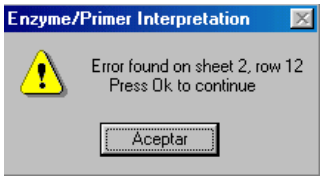


Figure 26. Genotypes appear after pressing OK in the enzyme/primer interpretation box.

If there are missing data for any of the individuals, *Transformer-3* will just assign a "999" to the corresponding genotype.

If you have drawn more than two alleles in an individual at the given locus, *Transformer-3* will warn you (see the error message below) so you can correct it before carrying on.



If you do not correct the mistake(s), the program will continue running, but you will probably generate defective files or data (see section 2). Therefore, you are strongly adviced to correct any mistake before passing on to interpret the next pattern.

Once you have completed this process for all the enzymes/primers included in your interpretation,

a) Click on the button "Genotype file" in the toolbar and then select the option "From the drawing" (Figure 27)

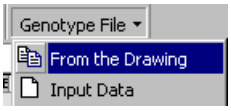


Figure 27

b) Select the loci for which you want to generate the genotype file by ticking on the appropriate boxes (default is all loci) [see Figure 28] and click "OK".

Transformer-3 will then generate a genotype sheet that is the basis for the subsequent calculations and data transformations. If you want to save this genotype workbook, do it now.

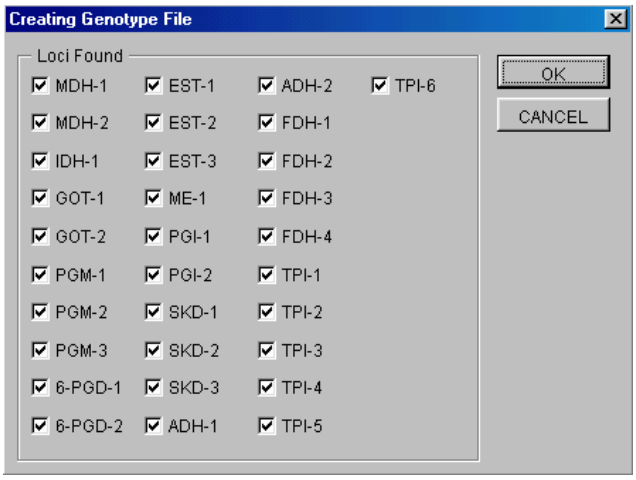


Figure 28. Box to select the loci you want to include in the analyses. By default, *Transformer-3* selects all loci in the project whose patterns have been interpreted (obtained from the file "transf-draw.xls")

1.2.11. Tips on the drawing utility

1. Check carefully the position and colour of the alleles before generating the corresponding genotype file. Remember that in order for an allele to be genotyped, its position and colour must correspond to one of those defined at the enzyme/primer header. Otherwise, *Transformer-3* will not consider it.

2. Take care not to draw more than two alleles per individual at a given locus. However, if you do so, *Transformer-3* will prompt an error message when you invoke the "Pattern Interpretation" command (see section 1.1.8). You have to correct the mistakes that *Transformer-3* will eventually pinpoint before moving on to the interpretation of the next locus.

3. The enzyme/primer patterns you interpret must be in the active sheet. If you introduce the code of an enzyme/primer that appears in another sheet of your project, *Transformer-3* will tell you that it cannot find that item in the current sheet.

4. If you want to change the position of one or several alleles after completing the process of pattern interpretation, you can do it, but you will have to press "pattern interpretation" for the affected loci and then "genotype file" again. Otherwise, the genotype file will be the same as the one without the change(s).

1.3. ENTERING A TABLE OF GENOTYPES

Users of *Transformer-3* that already have genotype matrices for their data may (rightly) consider that drawing their interpretations would be burdensome and time consuming. For such cases, *Transformer-3* offers the option of entering a table of genotypes.

1.3.1. How to input your genotypes for analysis

In the *Transformer-3* toolbar menu Select "Input data" on the option "Genotype file" (Figure 29)

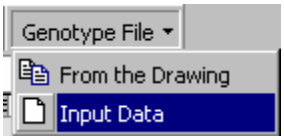


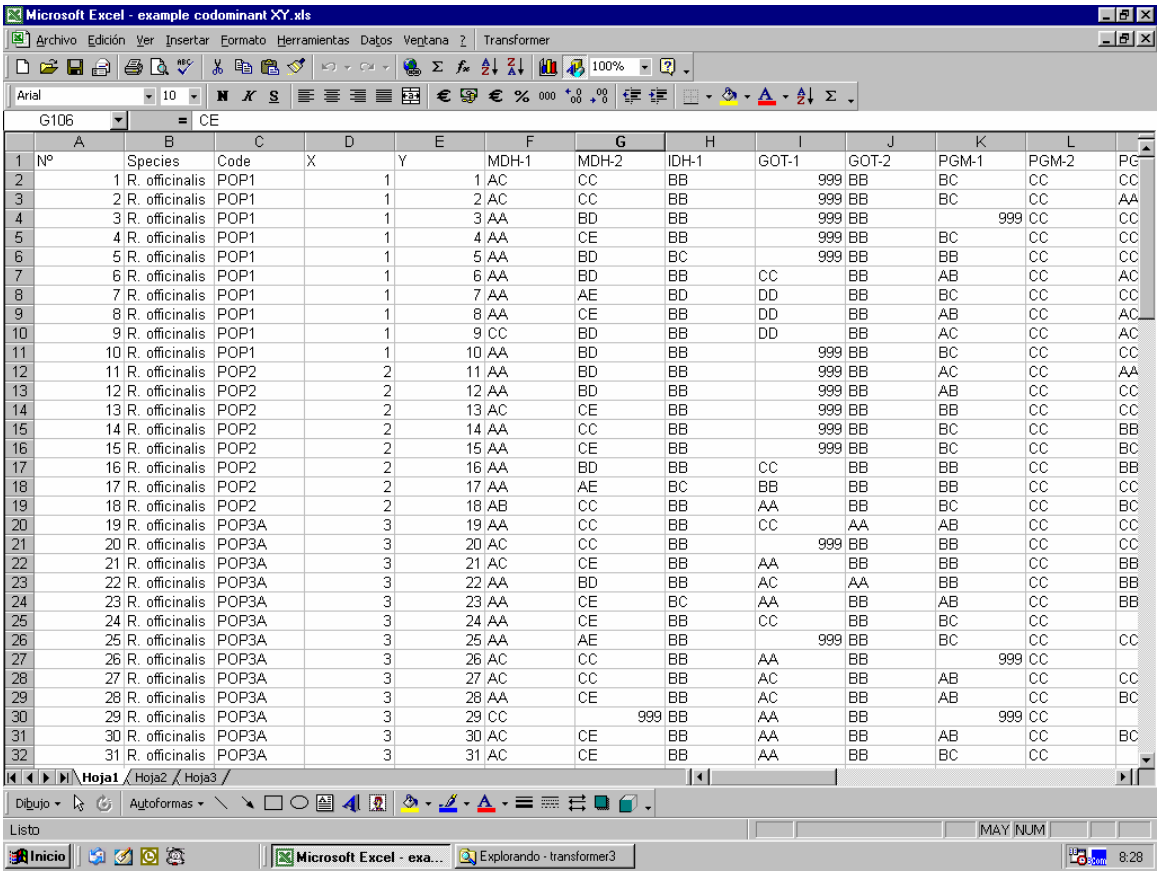
Figure 29. Selecting the option to enter a genotype Table in *Transformer-3*.

Then, you just need to have a Microsoft Excel genotype table like the one in the attached file "transf-gntp.xls" (see Figure 30 for an example). In this file,

a) The first line should contain the headers for the species, populations and loci.

   In this line, the first column is the sample number, the second is the species name, the third is the population code, and the fourth and the fifth (X and Y, respectively) are spatial coordinates to be used in the files for the programs SPAGEDI and Passage. If no spatial analysis is pursued with the data, then these columns can be left blank. Only make sure that the string of letters in the "population code" column is exactly the same for all the individuals that you want to include in a given population and follows the specifications in section 1.2.2.

b) The second line and the subsequent ones contain the data (**DO NOT** start writing your data in the first line).



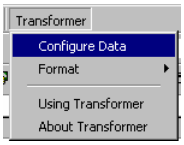
	A	B	C	D	E	F	G	H	I	J	K	L	
1	N°	Species	Code	X	Y	MDH-1	MDH-2	IDH-1	GOT-1	GOT-2	PGM-1	PGM-2	PG
2	1	R. officinalis	POP1	1	1	AC	CC	BB	999	BB	BC	CC	CC
3	2	R. officinalis	POP1	1	2	AC	CC	BB	999	BB	BC	CC	AA
4	3	R. officinalis	POP1	1	3	AA	BD	BB	999	BB	999	CC	CC
5	4	R. officinalis	POP1	1	4	AA	CE	BB	999	BB	BC	CC	CC
6	5	R. officinalis	POP1	1	5	AA	BD	BC	999	BB	BB	CC	CC
7	6	R. officinalis	POP1	1	6	AA	BD	BB	CC	BB	AB	CC	AC
8	7	R. officinalis	POP1	1	7	AA	AE	BD	DD	BB	BC	CC	CC
9	8	R. officinalis	POP1	1	8	AA	CE	BB	DD	BB	AB	CC	AC
10	9	R. officinalis	POP1	1	9	CC	BD	BB	DD	BB	AC	CC	AC
11	10	R. officinalis	POP1	1	10	AA	BD	BB	999	BB	BC	CC	CC
12	11	R. officinalis	POP2	2	11	AA	BD	BB	999	BB	AC	CC	AA
13	12	R. officinalis	POP2	2	12	AA	BD	BB	999	BB	AB	CC	CC
14	13	R. officinalis	POP2	2	13	AC	CE	BB	999	BB	BB	CC	CC
15	14	R. officinalis	POP2	2	14	AA	CC	BB	999	BB	BC	CC	BB
16	15	R. officinalis	POP2	2	15	AA	CE	BB	999	BB	BC	CC	BC
17	16	R. officinalis	POP2	2	16	AA	BD	BB	CC	BB	BB	CC	BB
18	17	R. officinalis	POP2	2	17	AA	AE	BC	BB	BB	BB	CC	CC
19	18	R. officinalis	POP2	2	18	AB	CC	BB	AA	BB	BC	CC	BC
20	19	R. officinalis	POP3A	3	19	AA	CC	BB	CC	AA	AB	CC	CC
21	20	R. officinalis	POP3A	3	20	AC	CC	BB	999	BB	BB	CC	CC
22	21	R. officinalis	POP3A	3	21	AC	CE	BB	AA	BB	BB	CC	BB
23	22	R. officinalis	POP3A	3	22	AA	BD	BB	AC	AA	BB	CC	BB
24	23	R. officinalis	POP3A	3	23	AA	CE	BC	AA	BB	AB	CC	BB
25	24	R. officinalis	POP3A	3	24	AA	CE	BB	CC	BB	BC	CC	
26	25	R. officinalis	POP3A	3	25	AA	AE	BB	999	BB	BC	CC	CC
27	26	R. officinalis	POP3A	3	26	AC	CC	BB	AA	BB	999	CC	
28	27	R. officinalis	POP3A	3	27	AC	CC	BB	AC	BB	AB	CC	CC
29	28	R. officinalis	POP3A	3	28	AA	CE	BB	AC	BB	AB	CC	BC
30	29	R. officinalis	POP3A	3	29	CC	999	BB	AA	BB	999	CC	
31	30	R. officinalis	POP3A	3	30	AC	CE	BB	AA	BB	AB	CC	BC
32	31	R. officinalis	POP3A	3	31	AC	CE	BB	AA	BB	BC	CC	

Figure 30. Detail of the format of a genotype sheet for entry in *Transformer-3*. The coordinates in columns X and Y are optional, and, if used, they are only considered for the formats of the programs SPAGEDI and Passage.

Section 2. Configuring the data

The starting point of the data configuration capabilities of *Transformer-3* is the matrix of genotypes, either obtained through the drawing sheet or implemented *ad hoc*. To configure your data for analysis:

1. Click on the button «Configure data» from the toolbar of *Transformer-3* (see below)



2. The big dialog box that will appear (see Figure 31) contains the default options. This box can already be used for data analysis and transformation (see section 2.1), but it also constitutes the basis to implement the different configurations that you may want to give to your data (see section 2.2). Before configuring the data, you have to specify whether they are dominant or co-dominant by ticking the appropriate box.

2.1. THE DEFAULT CONFIGURATION

The dialog box that first appears when you click on the option «Configure data» contains the configuration of data that *Transformer-3* would analyse by default (see Figure 31).

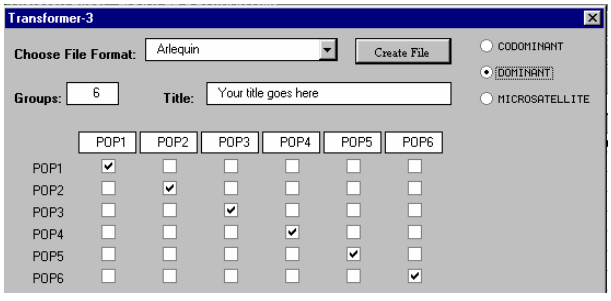


Figure 31. Default population configuration for six populations.

This box consists of three parts:

1. The number of groups with the default value, that corresponds to the number of different populations that *Transformer-3* has detected in the table of genotypes that the user has introduced (6 in Figure 31).

2. A field for choosing the file format(s) you want to transform your data into ("Choose File Format" bar). For any combination of populations, *Transformer-3* can give you the file formats needed to run your data in several population genetic programs, that vary depending on whether your data are codominant or dominant.

Furthermore, *Transformer-3* calculates the probabilities of loss (*L*) *sensu* Bengtsson, Weibull and Ghatnekar (1995) [see section 3.7] and a table of allele frequencies associated with the configuration of populations that you have defined (button "Prob. Loss")

3. A square matrix where the rows and columns are the populations that are included in the genotype table (6x6 in the example of Figure 31). Every cell in this matrix (there are 36 cells in the example) can be selected in order to define different configurations for analysis (see section 2.2). At present, the limit for the number of populations is 50.

4. A "Title" field where you can write the title of your data file.

If you choose any calculation option for this default configuration, the resulting analyses or data files that *Transformer-3* will generate will correspond to considering all the populations individually.

2.2. DEALING WITH GROUPS OF POPULATIONS

Many times, the population geneticist is interested in obtaining the values of the genetic polymorphism parameters for different groups in which the data can be subdivided. This utility of *Transformer-3* consists of four basic steps:

1) In the box "groups", choose the number of groups to be defined.

2) Press «Return»

   Then, the default matrix will reduce to a new one with the same number of rows as before but with the number of columns equalling the number of groups you defined.

3) Label the groups with proper names

4) Tick the cell(s) corresponding to the population(s) you want to include in each group.

In *Transformer-3*, a group can consist of any number of populations (one population can be a group), and one given population can appear in more than one group at the same time.

In the sections below, we discuss several possible options to define population groups.

2.2.1 Analysing population subsets

Suppose you have a data set for a large number of populations but you are only interested in analysing only a certain sub-group of populations within it. This is how to do it:

a) Write the number of groups you wish to establish in the corresponding cell and then press «Return».

   In the example below, suppose we want to analyse only the five populations POP1, POP2, POP4, POP5 and POP6. Therefore, we first have to write a «5» in the cell «Populations». After pressing «Return», the original 6x6 matrix has changed into a new 6x5 matrix (see Figure 32).

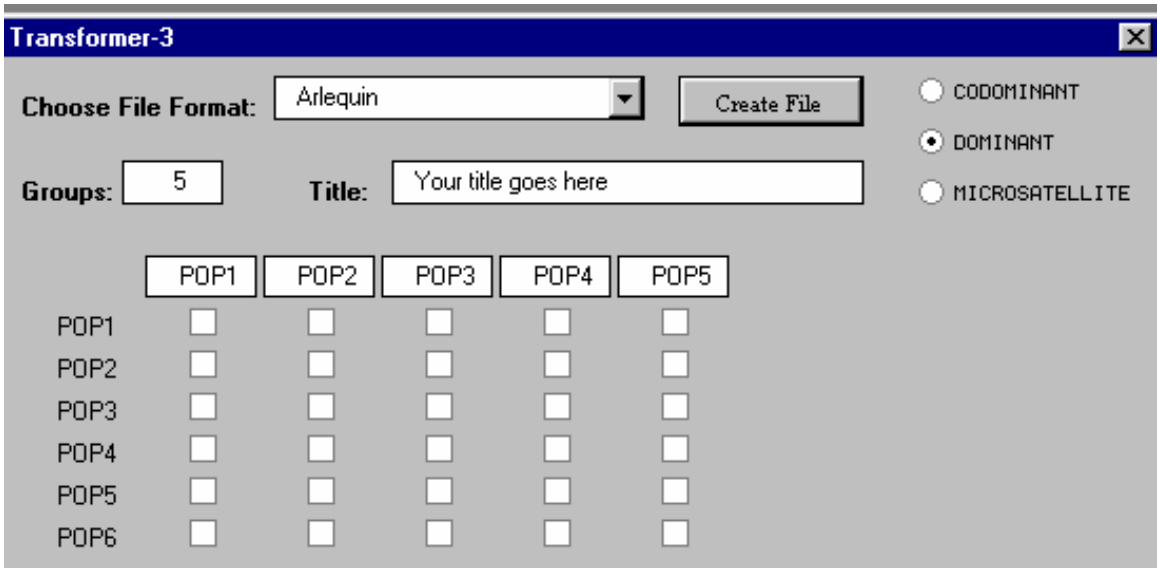


Figure 32.

b) Write the names of the populations you wish to include in this partial analysis in the column headers

In the example (Figure 33), we write these names in the cells above the columns of the matrix.

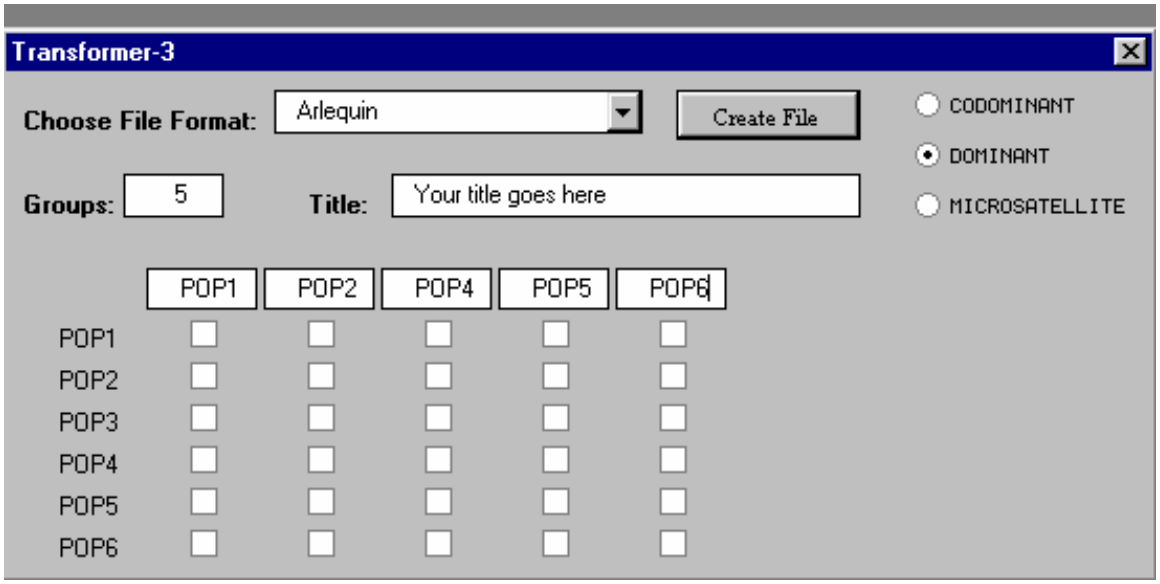


Figure 33.

b) Tick on the cells of the new matrix corresponding to the populations you want to analyse under this data configuration.

In the example, we have to tick only the cells that correspond to the populations we want to analyse (Figure 34)

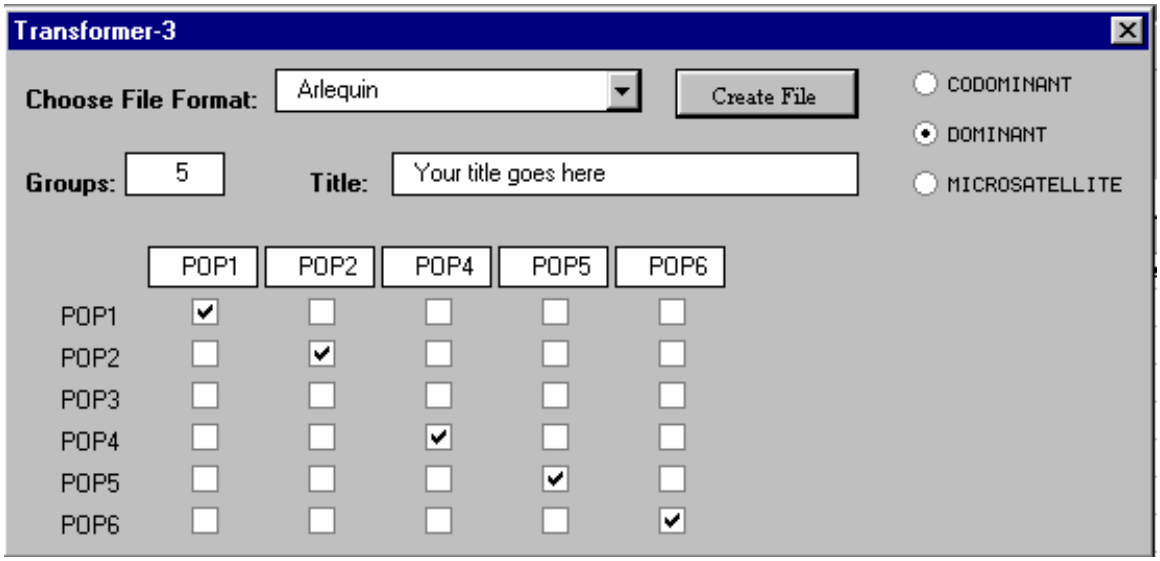


Figure 34.

c) Press the button corresponding to the transformation(s) you want to perform for that conformation of your data, give adequate names to the files that will be generated and press «save» (see section 3).

After this, *Transformer-3* is ready to perform the calculations and obtain the formats for that set (see section 3).

2.2.2. Comparing independent groups of populations

Most of the times, understanding the genetic relationships among the organisms we are analysing entails the comparison of groups defined using different criteria of interest (i. e., geographic distribution, specific ascription, habitat, clade ascription, etc…).

Transformer-3 allows the user to establish groups within the data in the following way:

a) Write the number of groups you wish to establish in the corresponding cell and then press «Return».

In the example in the file "transf-gntp.ex" (with 7 populations), imagine that populations POP1, POP2 and POP4 belong to a Species 1, and the remaining populations to a Species 2. If we wanted to compare these two species, we would write a «2» in the cell labelled «Groups» (Figure 35)

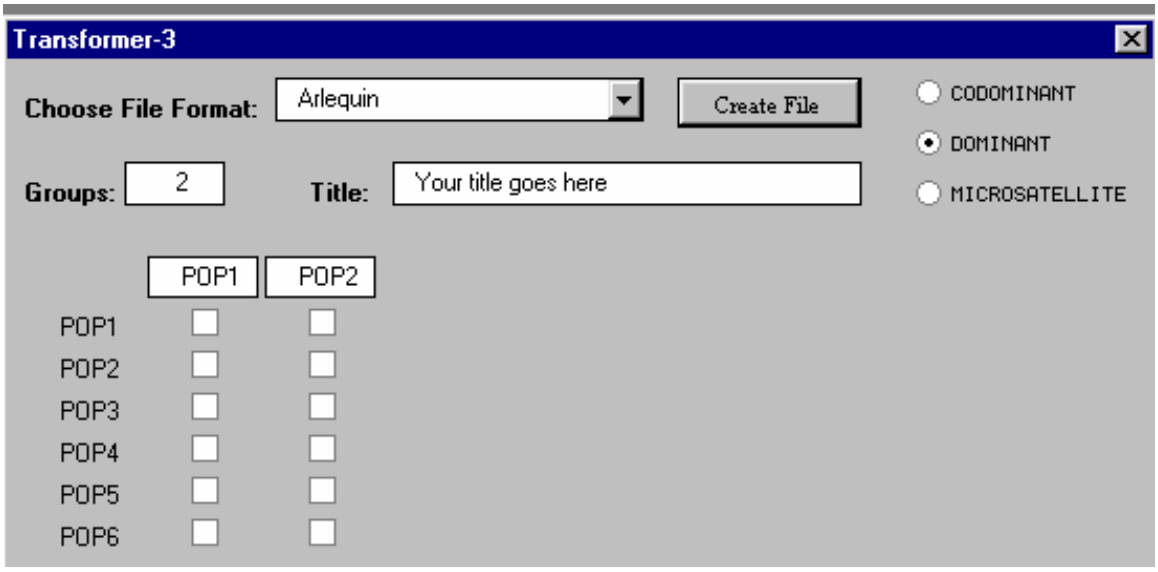


Figure 35.

b) Re-name the matrix column headers in the box to label the groups you want to define.

In our example, we choose the labels SP1 and SP2 (Figure 36).

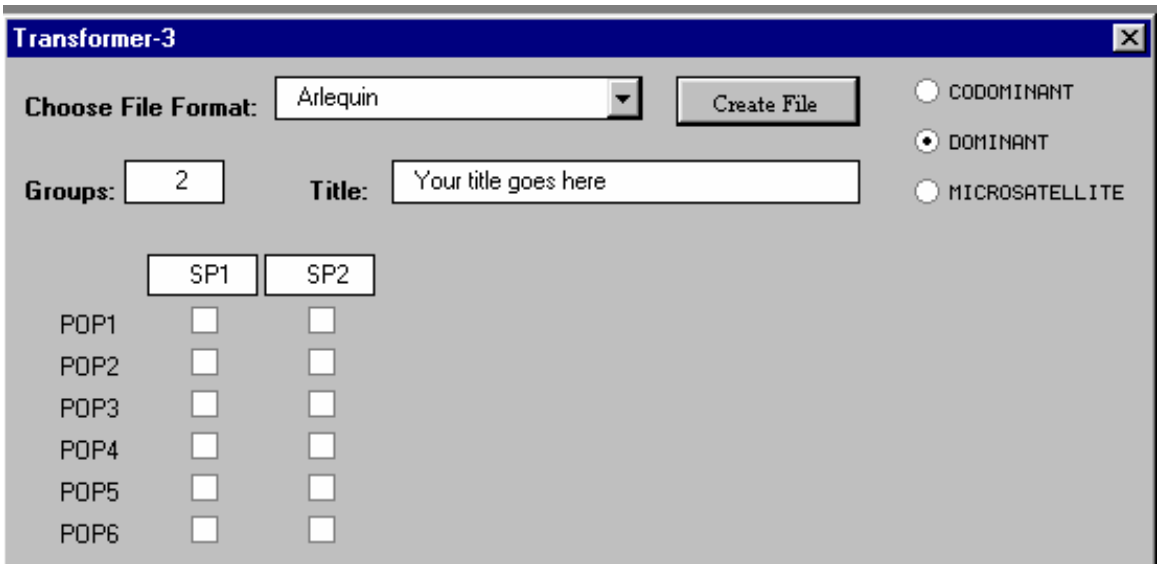


Figure 36.

c) For each of the groups, tick the boxes that correspond to the populations that they must contain.

In the example, group SP1 consists of the populations POP1, POP2 and POP4, while group SP2 consists of POP3, POP5 and POP6. Figure 37 illustrates the aspect that the configuration matrix would have in this case.

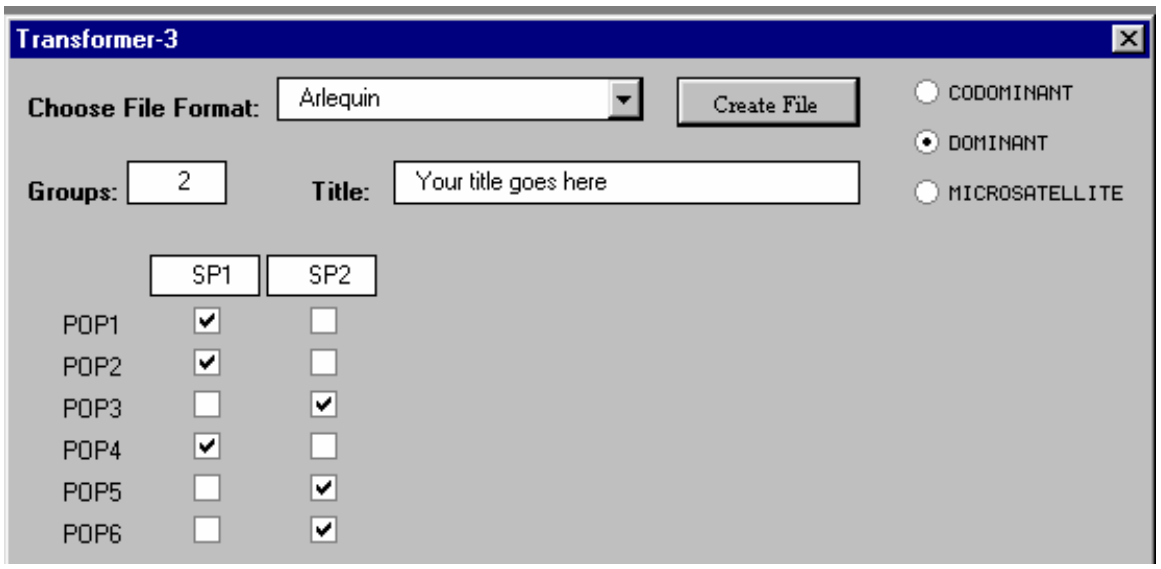


Figure 37.

d) Press the button corresponding to the transformation(s) you want to perform for that conformation of your data, give adequate names to the files that will be generated and press «save» (see section 3).

After completing this sequence, the files are ready to be run in the specific programs for which they were formatted (see section 3). If you have used the file "transf-gntp.xls" to follow this explanation, try some of the options.

2.2.3. Including populations in more than one group

In some cases, the population geneticist might be interested in testing how the values of different parameters change depending on which populations are included/removed from a given group. For these and other similar cases, *Transformer-3* allows the user to include any population in different groups in the following way:

1) Select the total number of groups to analyse

   In the example of Figure 38, imagine we want to define six groups with the data in the file "transf-gntp.xls".

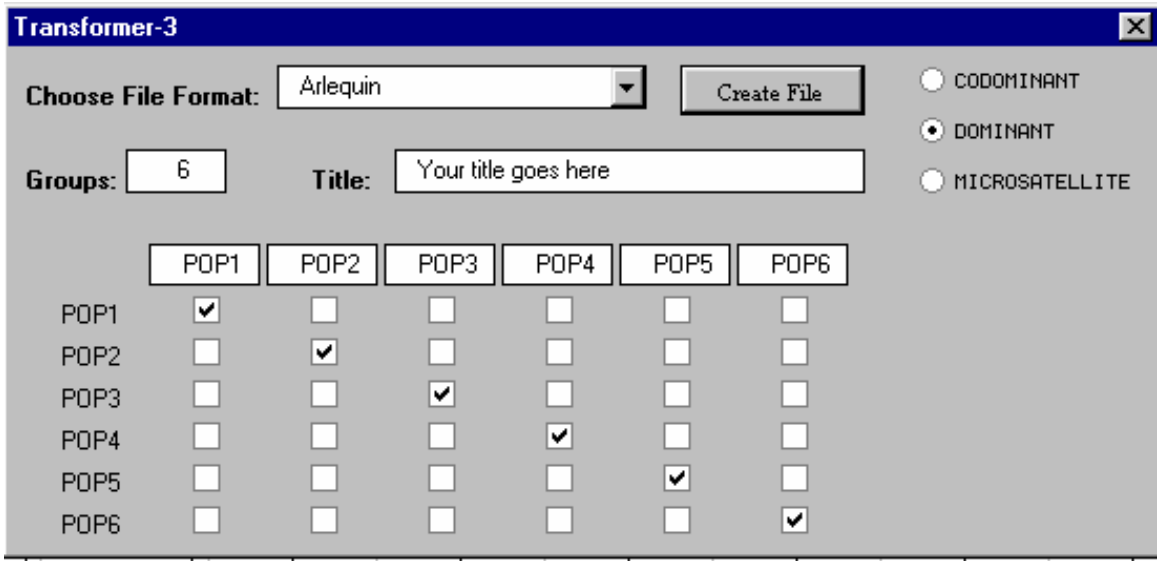


Figure 38.

2) Label the groups with a proper name

In Figure 39, we just choose the labels GR1 to GR6.

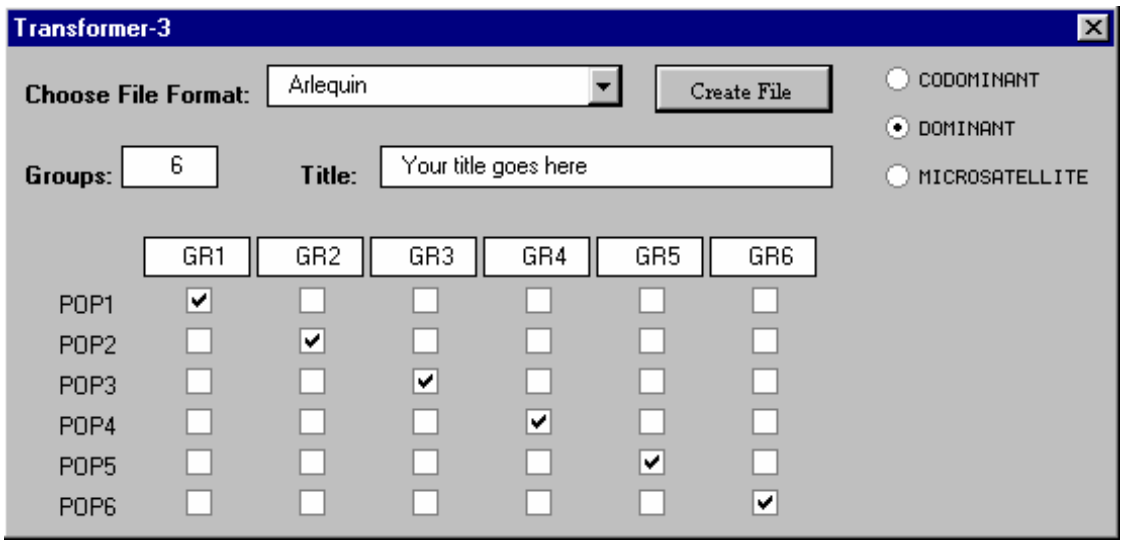


Figure 39

3) Select the populations to be included in each group

Imagine we want to check the effect of removing sequentially single populations in the value of some parameter that can be calculated using *Transformer-3*. The selections to make would be like in Figure 40.

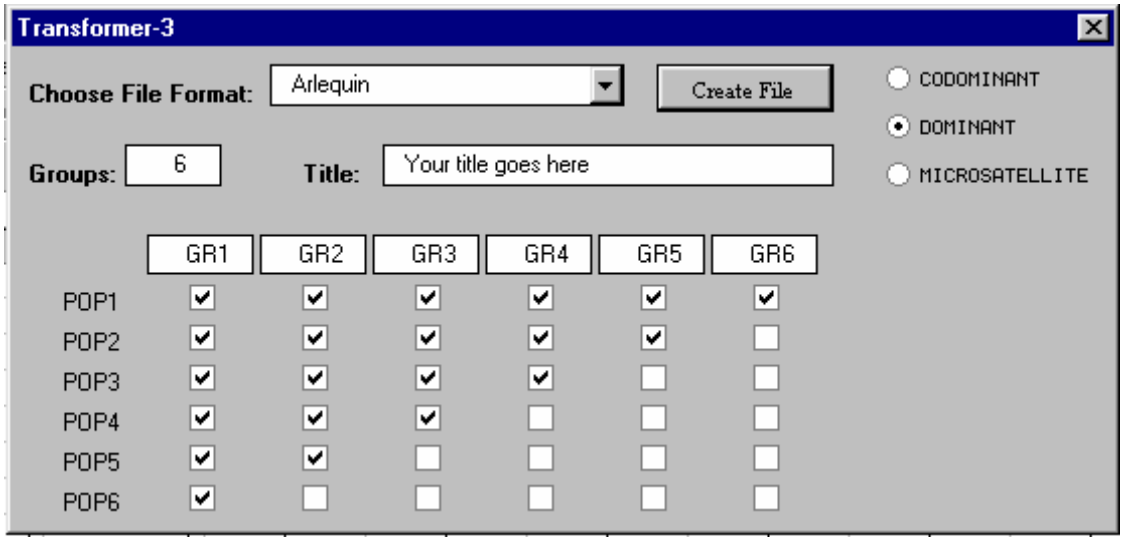


Figure 40

4) Press the button corresponding to the transformation(s) you want to perform for that conformation of your data, give adequate names to the files that will be generated and press «save» (see section 3).

After completing this sequence, the files are ready to be run in the specific programs for which they were formatted.

Section 3. Processing the data

For each configuration of populations, *Transformer-3* generates automatically the necessary files to run several population genetic analysis programs that vary slightly depending on whether the input data are dominant or co-dominant (Table 3.1), and (for co-dominant data only) it calculates all the parameters related to the probabilities of allelic loss.

Table 3.1 Present genotype transformation capabilities of *Transformer-3*, and references of the programs that it allows to work with at the time being. "Data" refers to whether *Transformer-3* deals with codominant (C) or dominant (D) genotype data matrices in each specific case.

Program	Data	Reference	Downloadable from
AFLPSurv	D	Vekemans (2002)	http://www.ulb.ac.be/sciences/lagev/aflp-surv.html
Arlequin	C, D	Schneider, Roessli and Excoffier (2000)	http://lgb.unige.ch/arlequin/
Biosys	C	Swofford and Selander (1989)	Get ordering info at swofford@csit.fsu.edu
Bottleneck	C	Piry, Luikart and Cornuet (1998)	http://www.montpellier.inra.fr/CBGP/softwares/
FSTAT	C	Goudet (1995)	http://www.unil.ch/izea/softwares/fstat.html
GDA	D	Lewis and Zaykin (1999)	http://alleyn.eeb.uconn.edu/gda/
GenAlex	D	Peakall and Smouse (2005)	http://www.anu.edu.au/BoZo/GenAlEx/
GenePop	C	Raymond and Rousset (1995)	http://wbiomed.curtin.edu.au/genepop/
GeneStat	C	Lewis and Whitkus (1993)	http://lewis.eeb.uconn.edu/lewishome/software.html
NTSys-pc	C, D	Rohlf (1988)	Order at http://www.exetersoftware.com/
PASSAGE	D	Rosenberg (2001)	http://www.passagesoftware.net/
PopGene	C, D	University of Alberta (1997)	http://www.ualberta.ca/~fyeh/
SPAGeDi	C, D	Hardy and Vekemans (2002)	http://www.ulb.ac.be/sciences/ecoevol/spagedi.html

3.1. THE BIOSYS FORMAT

Only codominant data

BIOSYS (Swofford and Selander 1989) is a Fortran IV computer program that can be used to calculate the values of most population genetic polymorphism indicators, test for Hardy-Weinberg equilibrium, compute *F*-statistics, perform heterogeneity chi-square analysis, calculate a variety of genetic distance coefficients, construct phenograms and estimate phylogenies through the distance Wagner procedure.

3.1.1. Obtaining the Biosys format

1) Select «Biosys» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

3) Your data are ready to run in Biosys.

By default, the ASCII file that *Transformer-3* creates for Biosys contains the following command lines at the end:

```
NEXT
END;
STEP VARIAB:
FULLOUT, PCRIT=2;
END;
STEP HDYWBG:
LEVENE, EXACTP;
END;
STEP SIMDIS:
ALLCOEF, SINGLE=2;
END;
STEP COEFOUT;
BELOW=1, ABOVE=2;
END;
STEP SINGLE:
COEF=1;
END;
STEP DISTRIB:
COEF=8;
END;
STEP FSTAT:
OUTPUT=1;
END;
STEP WRIGHT78:
END;
STEP WRIGHT78:
NOHRCHY;
END;
STEP HETXSQ:
CONTAB, SUBDIV=1;
END;
STEP CLUSTER:
COEF=1, COPHEN;
COEF=9;
END;
```

If you wish to remove commands or add new calculations, just do it removing or typing lines in this ASCII file.

3.2. THE BOTTLENECK FORMAT

Only codominant data

The program Bottleneck (Piry, Luikart and Cornuet 1998) applies a sign test for heterozygosity excess (Cornuet and Luikart 1996) to detect whether the populations have experienced recent historical bottlenecks. This test compares expected heterozygosity (H_e) under Hardy-Weinberg expectations to the heterozygosity expected at mutation-drift equilibrium (H_{eq}) in a sample that has the same size and the same number of alleles as the sample used to measure H_e (Luikart and Cornuet 1998). The rationale of the test is that, since low frequency alleles are lost at a much faster rate than heterozygosity in a bottleneck situation, bottlenecked populations are expected to have a heterozygote excess.

3.2.1. Obtaining the BOTTLENECK format

The Bottleneck option in *Transformer-3* gives you a single file that contains the format for all the populations or population groups in the configuration that you defined. To obtain it,

1) Select «Bottleneck» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

3) Your data are ready to run in Bottleneck

3.3. THE GENEPOP FORMAT

Only codominant data

GenePop (Raymond and Rousset 1995) is a software package that runs under the DOS operating system. The DOS version is updated periodically and contains a few options not available on the web site of the program (website: http://www.cefe.cnrs-mop.fr/). GenePop allows the user to perform most calculations and tests related to the estimation of population genetic variation from the information contained in molecular markers.

3.3.1. Obtaining the GenePop format

1) Select «GenePop» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

3) Your data are ready to run in GenePop

3.4. THE GeneStat FORMAT

Only codominant data

GeneStat-PC 3.31 (Lewis 1993) calculates polymorphism indices, gene diversities, genetic distances and Nei's (1973) population structure statistics (H_s, H_t, J_s and G_{st}).

3.4.1. Obtaining the GeneStat format

1) Select «GeneStat» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

3) Your data are ready to run in GeneStat

3.5. THE NTSYS FORMAT

Codominant and dominant data

Ntsys-pc 2.02j (Rohlf, 1998) is a multivariate statistical program that can be used for certain molecular population genetic data analyses. It consists of several different modules, and most procedures require the use of one or several of them. The most frequently used options in Ntsys by the population geneticists are the genetic distance calculations, clustering, multivariate analyses and Mantel tests.

3.5.1. Ntsys-PC format requirements

There are various entry formats in Ntsys. For allele frequencies, *Transformer-3* generates three files that Ntsys requires for this kind of data:

1. A data file with the allele frequencies
2. A sample size file
3. A locus size file

3.5.2. Obtaining the Ntsys format

Transformer-3 gives you the formats for the three different files required to run allele frequency data in Ntsys.

To obtain these files for any of your populations, do this:

1) Select «Ntsys» in the dialog box, and then "create file"

2) Give a proper name to each of the three files and save them

Transformer-3 reminds you what Ntsys file you are about to save. The first Ntsys file it creates is the allele frequency file (the "input file" for Ntsys), that appears in the dialog box with the default name "ntsys_frequencies" (Figure 41); just re-name the file as you wish.

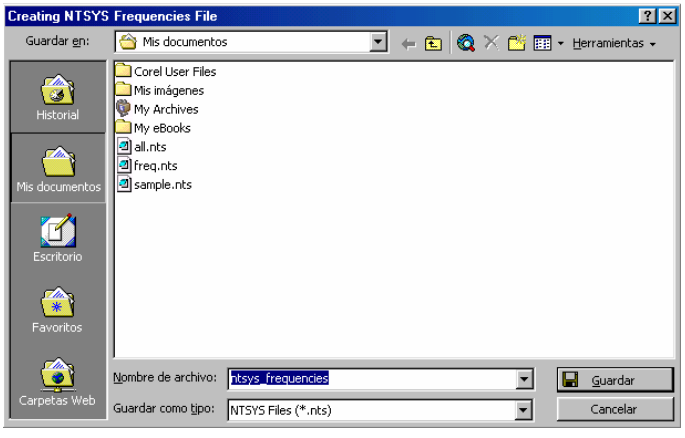


Figure 41.

Once you save this "input file", the second Ntsys file that *Transformer-3* will create for the configuration of populations you defined is the "loci array" file, that contains the number of alleles per locus. This file appears in the dialog box with the default name "ntsys_alleles" (Figure 42); just re-name it as you wish.

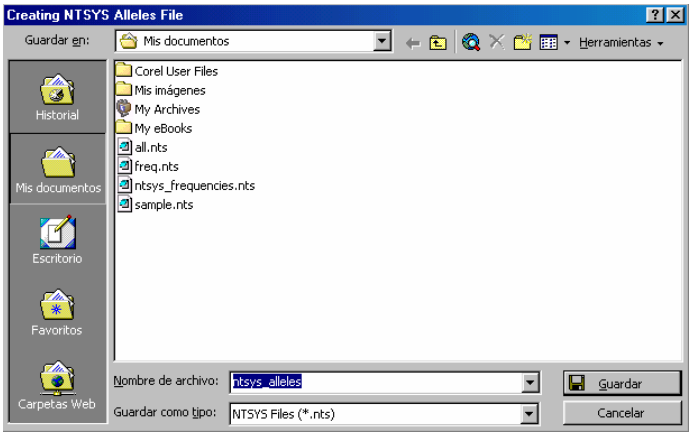


Figure 42

Finally, the third Ntsys file that *Transformer-3* creates for the configuration of data you defined is the "N array" file, that contains the corresponding sample sizes. This file appears in the dialog box with the default name "ntsys_samplesize" (Figure 43); just re-name it as you wish.

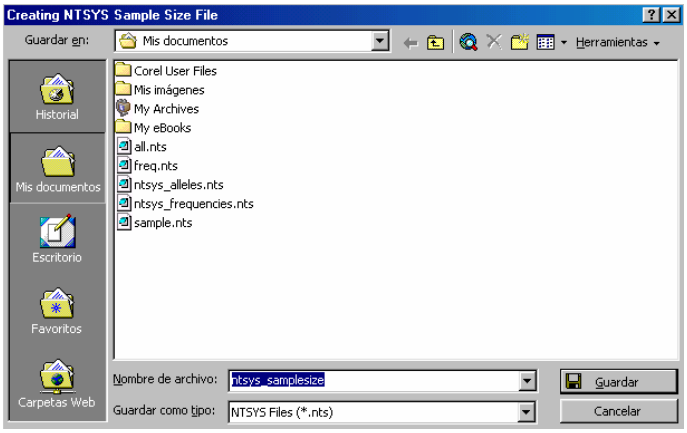


Figure 43.

3) Your data are ready to run in Ntsys

3.6. THE POPGENE FORMAT

Codominant and dominant data

PopGene version 1.32 (Yeh et al. 1997) is a program for the analysis of co-dominant and dominant diploid and haploid molecular markers.

It calculates most basic parameters of population genetic variation for different types of population structure and allows the user to perform many tests bearing on the structure of data (i.e., the homogeneity test, Ewens-Wattersson neutrality tests and the two-locus linkage disequilibrium test).

3.6.1. Obtaining the PopGene format

To obtain the PopGene file for any configuration of your populations, do this:

1) Select «PopGene» in the dialog box, and then "create file"

2) Give a proper name to the file and save it

3) Your data are ready to run in PopGene

3.7. THE AFLPsurv FORMAT

Only dominant data

AFLPsurv (Vekemans 2002) estimates genetic diversity and population genetic structure from population samples analyzed with AFLP or RAPD methods and computes genetic distance matrices.

3.7.1. Obtaining the AFLPsurv format

1) Select «AFLPsurv» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

3) Your data are ready to run in AFLPsurv

3.8. THE ARLEQUIN FORMAT

Codominant and dominant data

Arlequin (Schneider et al. 2000) is an exploratory population genetics software environment able to handle large samples of molecular data (RFLPs, DNA sequences, microsatellites), while retaining the capacity of analyzing conventional genetic data (standard multi-locus data or mere allele frequency data).

3.8.1. Obtaining the Arlequin format

1) Select «Arlequin» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

3) Your data are ready to run in Arlequin

3.9. THE FSTAT FORMAT

Only codominant data

FSTAT (Goudet 1995) is a computer package for PCs which estimates and tests gene diversities and differentiation statistics from codominant genetic markers. It computes both Nei and Weir & Cockerham families of estimators of gene diversities and F-statistics, and tests them using randomisation methods. Jackknife and Bootstrap confidence intervals are also provided.

3.9.1. Obtaining the FSTAT format

1) Select «FSTAT» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

3) Your data are ready to run in FSTAT

3.10. THE GDA FORMAT

Only dominant data

GDA (Lewis and Zaykin 1999) was designed to accompany the book "Genetic Data Analysis" by Bruce S. Weir (1996, Sinauer Associates). It computes linkage and Hardy-Weinberg disequilibrium, some genetic distances, and provides method-of-moments estimators for hierarchical F-statistics.

3.10.1. Obtaining the GDA format

1) Select «GDA» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

3) Your data are ready to run in GDA

3.11. THE GenAlex FORMAT

Only dominant data

GenAlex (Peakall and Smouse 2005) is a user-friendly cross-platform package for population genetic analysis that runs within Microsoft Excel(TM). GenAlEx enables population genetic data analysis of codominant, haploid and binary genetic data providing analysis tools applicable to plants, animals and microorganisms.

3.11.1. Obtaining the GenAlex format

1) Select «GenAlex» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

5) Your data are ready to run in GenAlex

3.12. THE PASSAGE FORMAT

Only dominant data

Passage (Rosenberg 2001) is a software package designed to aid in the analysis of data in a spatial context. This transformation capability is only operative in *Transformer-3* if your data source is a genotype matrix.

3.12.1. Obtaining the PASSAGE format

1) Select «PASSAGE» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

3) Your data are ready to run in PASSAGE

3.13. THE SPAGeDI FORMAT

Codominant and dominant data

SPAGeDI (Hardy and Vekemans 2002) is a computer package designed to characterize the spatial genetic structure of mapped individuals and/or mapped populations using genotype data of any ploidy level. This transformation capability is only operative in *Transformer-3* if your data source is a genotype matrix.

3.13.1. Obtaining the SPAGeDI format

1) Select «SPAGeDI» in the dialog box, and then "create file"

2) Give a proper name to the corresponding data file and save it

3) Your data are ready to run in SPAGeDI

3.14. THE PROBABILITIES OF ALLELIC LOSS

Only codominant data

Rare alleles are important in Conservation Biology because they represent unique evolutionary byproducts that may endow a species with advantageous properties to cope with eventual environmental shifts (Schonewald-Cox et al. 1983; Richter et al. 1994, Bengtsson et al. 1995, Caujapé-Castells 2004). Thus, collection designs oriented to sampling rare alleles provide the manager of genetic diversity with adequate tools with which to reinforce declining populations or aid the survival of reintroduced plants. The probability of allelic loss facilitates a straightforward way to analyse rare alleles and to incorporate them into conservation practice (Caujapé-Castells and Pedrola-Monfort 2004).

3.14.1. On the probability of loss

Transformer-3 calculates the probability of loss *L* (i. e., the probability that a sample of size *N* fails to include an allele with population frequency *p*) using the expression (Bengtsson et al. 1995)

$$L = (1 - p)^{2N}$$

Because these calculations are only suitable for alleles that are rare in some way (Bengtsson et al 1995), and there is no universally accepted definition of "rarity", *Transformer-3* offers two options to select the alleles for the calculation of *L*:

a) The default option, that we will call «Viera y Clavijo», follows Caujapé-Castells (2004), Caujapé-Castells and Pedrola-Monfort (2004) or Oliva et al. (2004) and calculates *L* only for the alleles that

   1) have an overall frequency ≤ 0.5, and

   2) are present in ≤ 50% of the populations considered,

b) The other option, that we will call "Select" enables the user to choose the alleles for these calculations by typing a "1" in the column labeled "Select" (Figure 44).

ALLELE	freq	N obs	Lo	Le	-log(Lo)	-log(Le)	Select
MDH-1A	0,7952	7	0	0	9,6404	9,6404	0
MDH-1B	0,059	2	0,784	0,4267	0,1057	0,3699	1
MDH-1C	0,188	7	0,0542	0,0542	1,266	1,266	0
MDH-2A	0,0596	5	0,541	0,4231	0,2668	0,3736	0
MDH-2B	0,1161	7	0,1777	0,1777	0,7502	0,7502	0
MDH-2C	0,461	7	0,0002	0,0002	3,7578	3,7578	0
MDH-2D	0,125	7	0,1542	0,1542	0,812	0,812	0
MDH-2E	0,2553	7	0,0161	0,0161	1,7925	1,7925	0
IDH-1A	0,0278	1	0,9452	0,6741	0,0245	0,1713	1
IDH-1B	0,8086	7	0	0	10,0532	10,0532	0
IDH-1C	0,0588	7	0,4278	0,4278	0,3688	0,3688	0
IDH-1D	0,3583	2	0,1695	0,002	0,7708	2,6977	0
IDH-1E	0,1833	1	0,6669	0,0587	0,1759	1,2314	0
GOT-1A	0,5187	6	0,0002	0	3,811	4,4461	0
GOT-1B	0,3748	4	0,0233	0,0014	1,6321	2,8561	0
GOT-1C	0,2341	7	0,0239	0,0239	1,6213	1,6213	0
GOT-1D	0,75	1	0,0625	0	1,2041	8,4288	0
GOT-2A	0,1554	4	0,2589	0,094	0,5868	1,0269	0
GOT-2B	0,9112	7	0	0	14,7221	14,7221	0
PGM-1A	0,1667	3	0,3349	0,0779	0,4751	1,1085	0
PGM-1B	0,3817	5	0,0082	0,0012	2,0877	2,9228	0
PGM-1C	0,4531	7	0,0002	0,0002	3,6698	3,6698	0
PGM-1D	0,3509	4	0,0315	0,0024	1,5015	2,6277	0
PGM-1E	0,0161	1	0,968	0,7964	0,0141	0,0989	1
PGM-2A	0,0459	3	0,7541	0,5176	0,1226	0,286	1

Figure 44: After choosing the option "Prob. Loss", a table like this one appears below the Table of allele frequencies. By default, the alleles selected are only the ones that fulfill the conditions described above under the name "Viera y Clavijo". However, you can make your own selections by typing 1 on the column "Select". In this Table, "freq" is the average allele frequency in the chosen group, "Nobs" is the number of populations where the allele was detected, and "Lo" and "Le" are the observed and expected probabilities of loss.

By default, *Transformer-3* obtains the value of both the expected probabilities of loss, L_e (i. e., assuming that the allele had its overall average frequency at each of the populations considered), and the observed probabilities of loss, L_o for all the alleles that fulfill the "Viera y Clavijo" conditions.

If you press the button "Create chart" that appears below the Table of the probabilities of loss (Figure 45), the values of L_o and L_e are used for two linear regression analyses (Bengtsson et al. 1995) where the average frequency of each allele is the *x*-axis and -log L_o and -log L_e are the respective *y*-axes (Figure 46). The chart is created only for the alleles that are selected in the column labeled "Selection".

TPI-5B	0,961	7	0	0	19,7229	19,7229	0
TPI-5C	0,046	2	0,8285	0,5176	0,0817	0,286	1
TPI-6A	0,3261	5	0,0193	0,004	1,7137	2,3992	0
TPI-6B	0,7671	7	0	0	8,8598	8,8598	0
		CREATE CHART					

Figure 45. The button "Create chart" appears at the end of the table that contains the probabilities of loss.

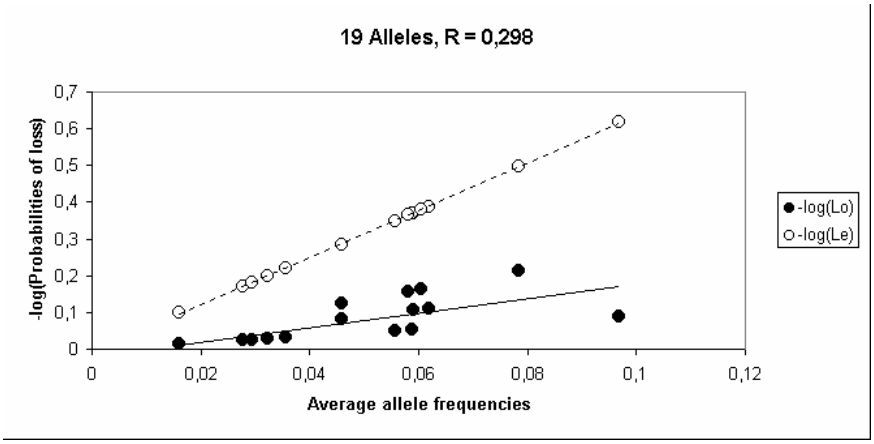


Figure 46. Linear regression of the average allele frequencies and the –log(Lo) [black circles, continuous line] and –log(Le) [black circles, discontinuous line] for all the alleles that fulfilled the "Viera y Clavijo" conditions in the example file "transf-gntp.xls". At the header of the chart, the program indicates the number of alleles that are included in the representation and calculates the representativity value as described in the text.

Transformer-3 also calculates the value for the representativity (*R*) of sampling only one population of that group relative to the total sample of rare alleles by dividing the slope of the observed regression line (based on the values of L_o) by the slope of the expected regression line (based on the values of L_e) [Bengtsson et al. 1995].

3.14.2. Obtaining the probabilities of allelic loss

1) Select «Prob. Loss» in the dialog box, and then "create file"

2) Select the option that you want to use (see section 3.14.1)

3) Give a proper name to the Excel output file and save it

The resulting Excel file will contain:

a) a table of allele frequencies for each of the groups selected,

b) a table with the values of *L* for the alleles that fulfilled the conditions of the calculation option that you selected (below the previous one),

c) the graph with the linear regressions commented in section 2.3.14.1. (if you pressed "Create Chart"), and

d) the value of representativity *R* (if you pressed "Create Chart").

3.15 TIPS ON PROCESSING THE DATA

After obtaining the datafiles and analyses for a given configuration of populations, you can go back to the original matrix and define another configuration of interest. This way you can get all files you want to analyse for as many configurations you may be interested in before running the corresponding programs or carrying out calculations.

LITERATURE CITED

Arús P, Shields CR (1983) Cole crops (*Brassica oleracea* L.) Isozymes in Plant Genetics and Breeding, Part B. Elsevier Publishers, Amsterdam.

Bengtsson BO, Weibull P, Ghatnekar L (1995) The loss of alleles by sampling: a study of the common outbreeding grass *Festuca ovina* over three geographical scales. *Hereditas* **122**, 221-238.

Caujapé-Castells J (2001) *Transformer-1*: Un programa de formateado de datos genotípicos individuales para implementación en programas de análisis molecular en genética de poblaciones. Distributed by the author.

Caujapé-Castells J, Baccarani-Rosas M (2004). *Transformer-2*: A program for the análisis of molecular population genetic data. Jardín Botánico Canario "Viera y Clavijo" & Instituto Tecnológico de Canarias. (http://www.jardincanario.org).

Caujapé-Castells J, Pedrola-Monfort J (2004) A sampling design for the ex-situ genetic conservation of the Ibero-Moroccan endangered endemic *Androcymbium gramineum*: implications for the assessment of a conservation strategy from a survey of genetic diversity for neutral markers. *Conservation Genetics* **5**, 131-144.

Caujapé-Castells J (2004) Manual para el muestreo genético de poblaciones naturales de plantas vasculares del Jardín Botánico Canario "Viera y Clavijo". Publicaciones EXEGEN.

Cornuet JM, Luikart G (1996) Description and evaluation of two tests for detecting recent bottlenecks. *Genetics* **144**, 2001-2014.

Goudet J (1995) Fstat version 1.2: a computer program to calculate F-statistics. *Journal of Heredity* **86**(6), 485-486.

Hardy OJ, Vekemans X (2002) SPAGeDi : a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.

Lewis PO (1993) *GeneStat-PC 3.31*. North Carolina State University. Raleigh. North Carolina.

Lewis PO, Zaykin D (1999) Genetic data analysis. Computer program for the analysis of allelic data. Version 1.0 (d12). Distributed by the authors at http://lewis.eeb.uconn.edu/lewishome/ software.html. Lewis Labs, Univ. of Connecticut, Storrs, CT.

Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* **12**, 228-237.

Oliva-Tejera F, Caujapé-Castells J, Naranjo-Suárez J, Navarro-Déniz J, Acebes-Ginovés JR, Bramwell D (2004) Variación genética de los *Lotus* L. (Fabaceae: Loteae) de pinar de Gran Canaria. *Botánica Macaronésica* **25**, 31-52.

Peakall R, Smouse PE (2005) GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. The Australian National University, Canberra, Australia.

Piry S, Luikart G, Cornuet JM (1998) Bottleneck, a program for detecting recent effective population size reductions from allele frequency data. INRA, URLB, Laboratoire de Modélisation et Biologie Evolutive. Montpellier, France.

Raymond M, Rousset F (1995) GenePop (version 1.2) population genetic software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249. (website: http://www.cefe.cnrs-mop.fr/).

Richter TS, Soltis PS, and Soltis DE (1994) Genetic variation within and among populations of the narrow endemic, *Delphinium viridescens* (Ranunculaceae). *American Journal of Botany* **81**, 1070-1076.

Rohlf FJ (1998) NTSYS-pc. Numerical Taxonomy and Multivariate Analysis version 2.02j. Applied Biostatistics Inc.

Rosenberg MS (2001) *PASSAGE.* Pattern analysis, spatial statistics and geographic exegesis. Version 1.0. Department of Biology, Arizona State University, Tempe, AZ.

Schneider S, Roessli D, Excoffier L (2000) *Arlequin: A software for population genetic data*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schonewald-Cox CM, Chambers SM, MacBride B, Thomas WL, editors (1983) *Genetics and conservation: a reference for managing wild animal and plant populations* Benjamin Cummings, Menlo Park, CA.

Swofford DL, Selander RB (1989) BIOSYS-1: a computer program for the analysis of allelic variation in genetics. University of Illinois, Urbana, Ill.

Vekemans X (2002) AFLPsurv version 1.0: a program for genetic diversity analysis with AFLP (and RAPD) population data. Laboratoire de Génétique et Ecologie végétales, Université Libre de Bruxelles.

Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX (1997) PopGene, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada.

APPENDIX: THE EXAMPLE FILES

Transformer-3 comes with two interrelated example files for codominant (allozyme) data and with one example file for dominant data.

Use these files to make your first trials with *Transformer-3* and to check the versatility of the program. You may want to add new enzymes, loci, primers, alleles or individuals, and then try to perform calculations and generate data files for all the configurations of loci and populations you may think of. This way, you will get acquainted with the program before you input your own real data.

The example files for co-dominant data

One of the two co-dominant files supplied is an example of drawn interpretations ("transf-draw.xls"), and the other one is a genotype file ("transf-gntp.xls") that corresponds to all loci that appear in the drawn interpretations (you can obtain it by interpreting the patterns and then pressing "Genotype file", as described in this manual).

Both files consist of data for 12 allozymes (31 loci) in 116 individuals that represent seven populations of the fictitious species *R. officinalis*. The drawings of these enzymes are distributed in two sheets within the same Excel file. Sheet 1 contains the patterns for MDH, IDH, GOT, PGM, 6-PGD, EST, ME and PGI, while sheet 2 contains the patterns for SKD, ADH, FDH and TPI. Rather than to provide the user with a real case, these examples try to account for a panoply of possible situations that the population geneticist might be confronted with when analysing molecular patterns for diploid codominant markers.

Although most of the drawings are based on real patterns obtained for different Canarian endemics at the Laboratorio de Biodiversidad Molecular of the Jardín Botánico Canario "Viera y Clavijo", these examples do not correspond to any real organism. They also incorporate several locus configurations that were drawn on purpose to illustrate how *Transformer-3* deals with particularly complex situations.

There are enzymes with just one associated locus and enzymes with many associated loci, monomorphic loci (ME-1 and PGI-1),

moderately polymorphic loci, and extremely polymorphic loci. The patterns for the enzymes also contain missing data and phantom bands for different individuals at different loci, and consider allele positions that would be particularly error-inducing if the molecular patterns were interpreted (or corrected) by hand.

The example file for dominant data

The example of dominant data represents six populations of the fictitious species *R. officinalis* for which 65 RAPD loci have been scored.