OligoSpecificitySystem

Version 1.0

Specificity analysis of oligonucleotide sets taking into account degeneracy and the mismatch possibilities

R.J. Michelland, S. Combes and L. Cauquil

Contact: laurent.cauquil@toulouse.inra.fr

Welcome to OLIGOSPECIFICITYSYSTEM.

OLIGOSPECIFICITYSYSTEM program is a free package for the R statistical program (R development Core Team, 2009) with a user-friendly graphical interface (GUI). It is intended to help scientists to analyse or to design optimal systems of oligonucleotides used in technologie based on PCR and nucleic acids hybridisation (PCR, real-time PCR, micro-arrays etc.). OLIGOSPECIFICITYSYSTEM is able to calculate global theoretical matching efficiency whatever the complexity of oligonucleotide system: number of oligonucleotides, oligonucleotide with degenerate sites and mismatch occurrence. Even if it has been specifically written for ARB program output (Ludwig *et al.* 2004), OLIGOSPECIFICITYSYSTEM allows any ASCII files (text files readable by common editors) to be imported thanks to the possibility of specifying the symbol used as separator between sequences in the ASCII file.

For advanced use, all procedures can be executed from the R prompt.

My colleagues and I will be happy to help you if you encounter any problems (laurent.cauquil@gmail.com).

CONTENTS

1- Introduction to OligoSpecificitySystem

1-1 Installation

The OLIGOSPECIFICITYSYSTEM program and R work on a wide variety of UNIX, Windows and MacOS platforms. Before installing OLIGOSPECIFICITYSYSTEM you first need to install R (R development Core Team, 2009). Sources, binaries and documentation for R can be obtained via CRAN, the "Comprehensive R Archive Network" (http://cran.rproject.org/mirrors.html). To install the OLIGOSPECIFICITYSYSTEM program, write at the R prompt:

> install.packages("OligoSpecificitySystem",dependencies=T)

Select the CRAN mirror and the OLIGOSPECIFICITYSYSTEM package. Then the package and its dependencies will be downloaded and installed.

1-2 Running

Once you start R, you don't need anymore to re-install the OLIGOSPECIFICITYSYSTEM program but you need to load it by writing at the prompt:

> library(OligoSpecificitySystem)

Once loaded, to run OLIGOSPECIFICITYSYSTEM you can write at the prompt one of the following line command:

> OligoSpecificitySystem()

or

> oligospecificitysystem()

or

> Oligospecificitysystem()

or

> OSS()

or

> Oss()

or

> oss()

1-3 Presentation of the main window

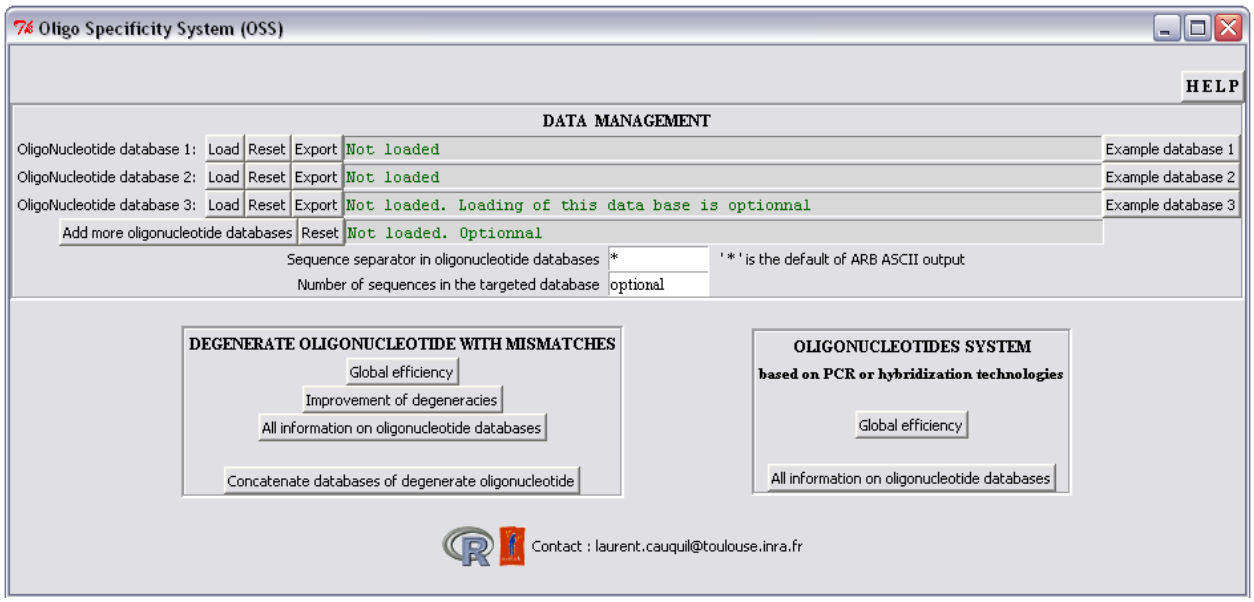


Figure 1. The main window of the program

The main window of OLIGOSPECIFICITYSYSTEM is divided in 4 sections (figure 1):

- The HELP button permits to access to the user guide manual.
- The DATA MANAGEMENT section permits to load, reset and export oligonucleotide databases. This section also permits to specify the sequence separator and optionnaly the number of sequences in the targeted database.
- The DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES section permits to compute the efficiency of the degenerate oligonucelotide set system, to evaluate the improvement of each degeneracy and to concatenate each database of degeneracy into a single database.
- The OLIGONUCLEOTIDES SYSTEM section permits to compute the efficiency of a oligonucleotide set like PCR-based or hybridization-based technologies.

2- Data management

The first step consists in importing oligonucleotide databases using the LOAD buttons. They open a window where you can select the file containing a database. If you have more than 3 databases, you can use the ADD MORE OLIGONUCLEOTIDE DATABASES button to select a folder containing more databases in separated ASCII files. Once loaded, the text bar indicates the path of the file loaded and the file name.

You can also easely test OLIGOSPECIFICITYSYSTEM program by opening examples of oligonucleotide database using the EXAMPLE DATABASE buttons. These will load databases in the program and open them in a text editor to see how to encode databases in ASCII files.

Two others functions in the data management section are the RESET and the EXPORT buttons. They allow respectively to reset and to export databases in ASCII. Notice that the LOAD buttons will automatically reset the current database before loading a new one.

The second step consists in specifying the species separator of the oligonucleotide databases. The easier way to determine the species separator is to open one database in a text editor before (figure 2). In OLIGOSPECIFICITYSYSTEM program, the species separator is "*" by default because it is the species separator of ARB program (Ludwig et al, 2004) output. The optional third step consists to specify the number of sequences in the targeted database. By default, all results are expressed in term of number of sequences. If the number of sequences in the targeted database is specify, all results are expressed in term of percentage of sequence compared to the number of sequences in the targeted database.

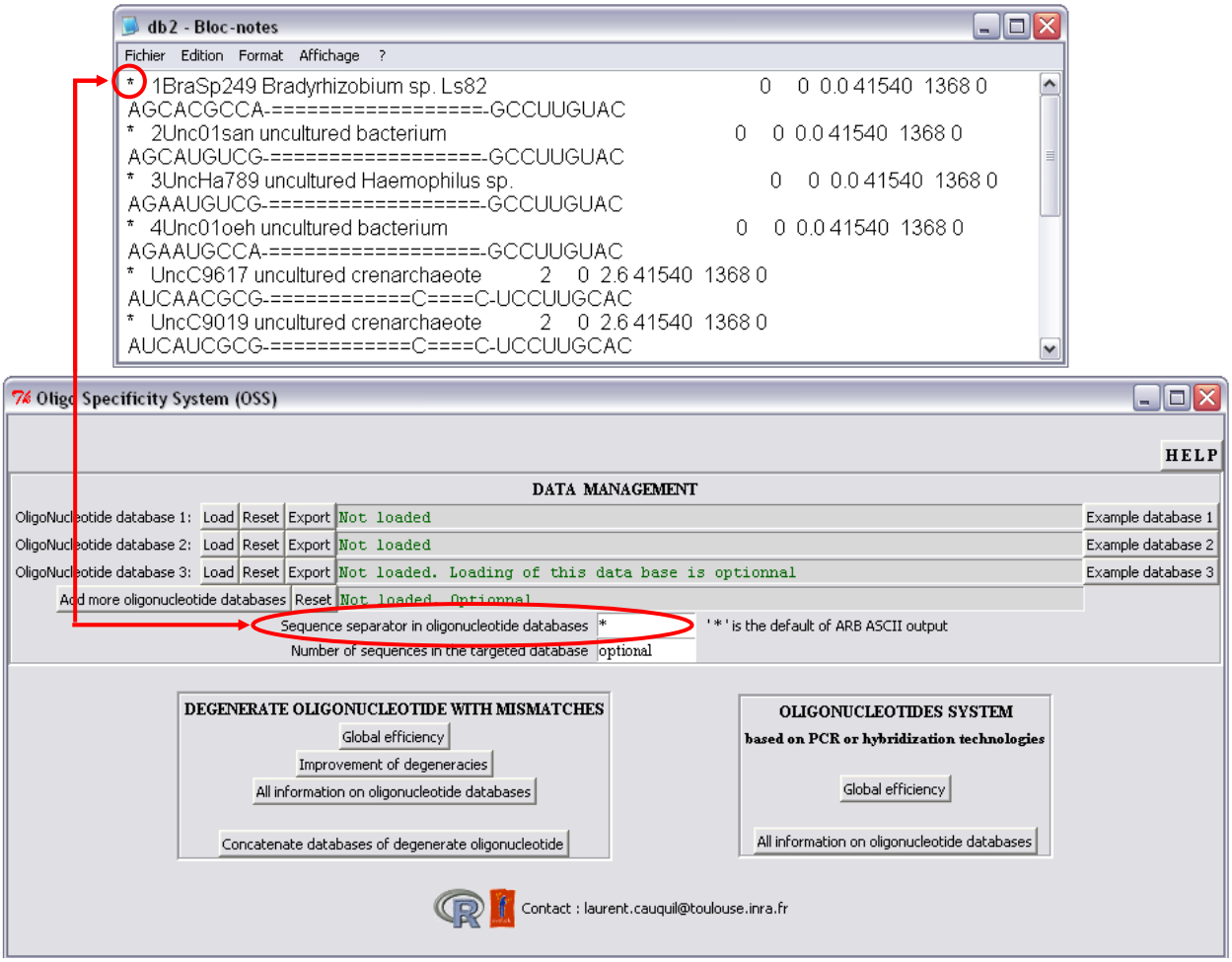


Figure 2. Specify the species separator in the oligonucleotide databases

3- Degenerate oligonucleotide with mismatches

The DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES section of OLIGOSPECIFICITYSYSTEM program allows to:

- compute the global efficiency of the studied oligonucleotide with or without mismatches with 2 databases (Figure 3A), 3 databases (Figure 3B) or more (not shown) using the GLOBAL EFFICIENCY button.
- evaluate improvements of degeneracies with 2 databases (Figure 3C) or 3 databases (Figure 3D) using the IMPROVEMENT OF DEGENERACIES button. This function is not available for more than 3 databases.
- show all information on databases of the studied oligonucleotide with 2 databases (Figure 3E) or 3 databases (Figure 3F) using the ALL INFORMATION ON OLIGONUCLEOTIDE DATABASES button. This function is not available for more than 3 databases.

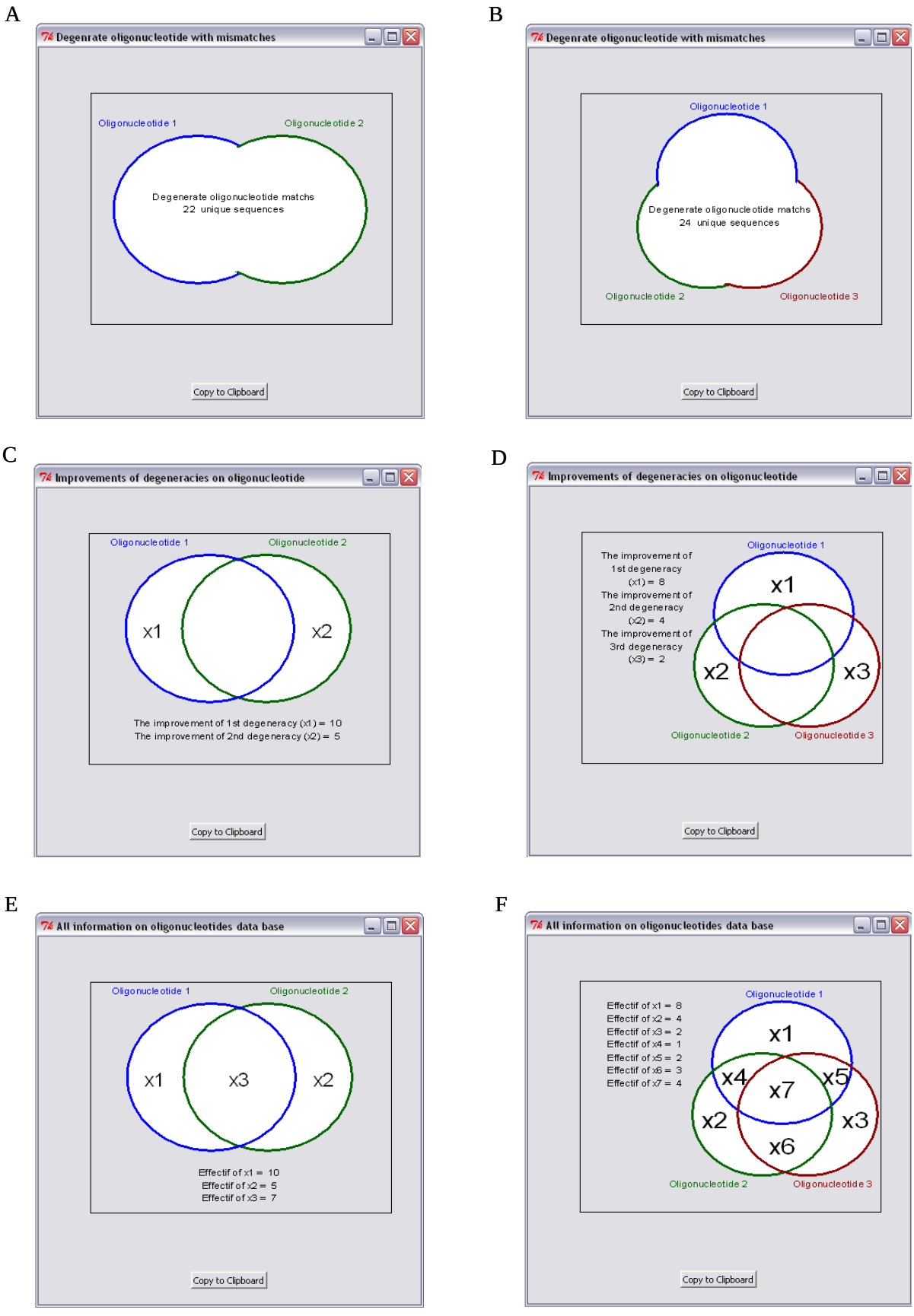


Figure 3. Resulted windows of GLOBAL EFFICIENCY (A, B), IMPROVEMENT OF DEGENERACIES (C, D) and ALL INFORMATION ON OLIGONUCLEOTIDE DATABASES (E, F) procedures in the DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES section

The function CONCATENATE DATABASES OF DEGENERATE OLIGONUCLEOTIDE of the DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES section (Figure 4) permits to concatenate all databases of a degenerate oligonucleotide in a single database. The created single degenerate oligonucleotide database can be either exported or loaded directely in one of the 3 oligonucleotide databases of the program.

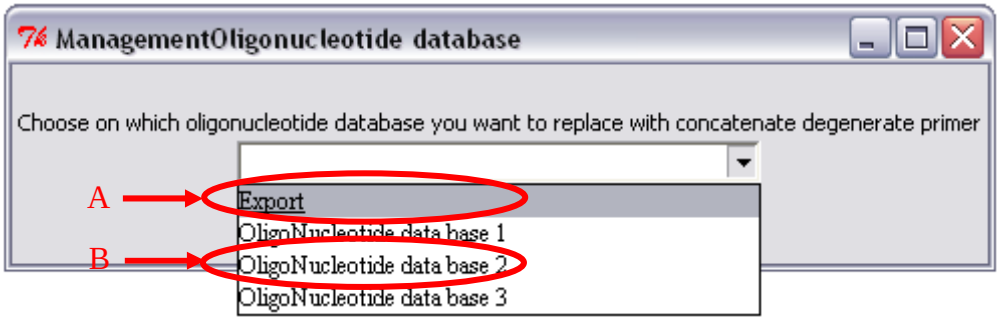


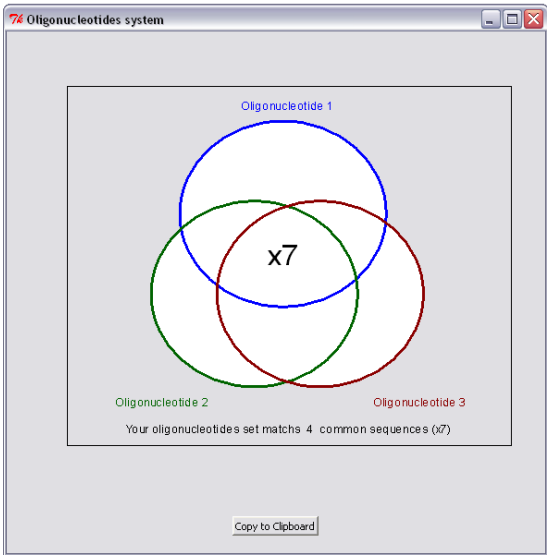
Figure 4. Window of the CONCATENATE DATABASES OF DEGENERATE OLIGONUCLEOTIDE procedure in the DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES section

4- Oligonucleotides system

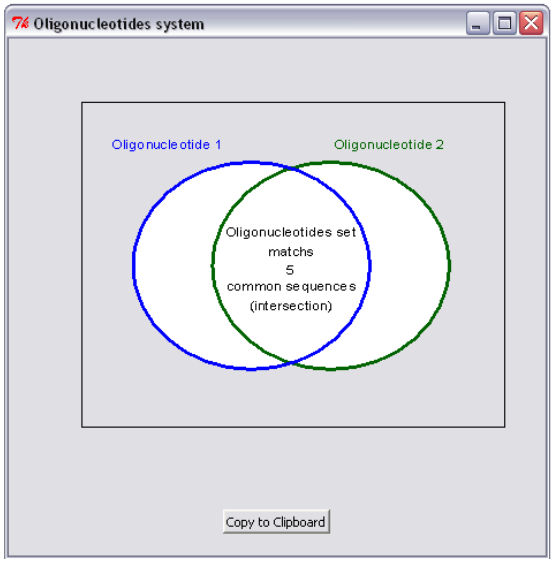
The OLIGONUCLEOTIDES SYSTEM section of OLIGOSPECIFICITYSYSTEM program allows to:

- compute the global efficiency of the oligonucleotide system with 2 databases (fFigure 5A), 3 databases (Figure 5B) or more databases (not shown) using the GLOBAL EFFICIENCY button.
- calcule all the information on databases of the oligonucleotide set with 2 databases (Figure 5C) or 3 databases (Figure 5D) using the ALL INFORMATION ON OLIGONUCLEOTIDE DATABASES button. This function is not available for more than 3 databases.

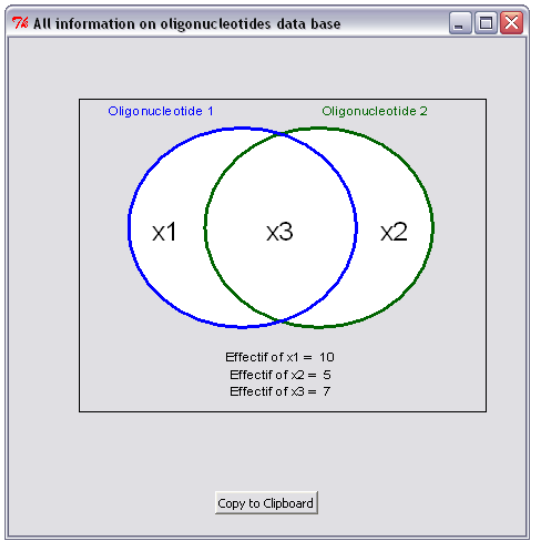
A



B



C



D

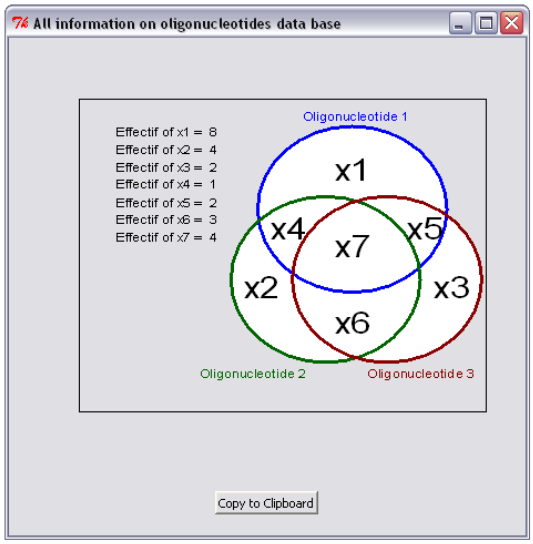


Figure 5. Resulted windows of GLOBAL EFFICIENCY (A, B) and ALL INFORMATION ON OLIGONUCLEOTIDE DATABASES (C, D) procedures in the OLIGONUCLEOTIDES SYSTEM section

5- Advanced mode: involved objects, involved procedures and principles of calculation

5-1 Objects invoked

Here is a list of the 9 objects used to run a project with OLIGOSPECIFICITYSYSTEM program:

- ONDB1name: this object is the file directory where the ACSI file of your 1st oligonucleotide database loaded is located.
- ONDB2name: this object is the file directory where the ACSI file of your 2nd oligonucleotide database loaded is located.
- ONDB3name: this object is the file directory where the ACSI file of your 3rd oligonucleotide database loaded is located.
- ONDB4name: this object is the file directory where the loaded folder containing ACSI files of your 4th oligonucleotide database is located.
- txt1: this object contains the 1st oligonucleotide database.
- txt2: this object contains the 2nd oligonucleotide database.
- txt3: this object contains the 3rd oligonucleotide database.
- txt4: this object contains the 4th oligonucleotide database.
- Separator: this object contains the species separator in oligonucleotide databases, e.g. "*" for ASCII files generated with ARB program.

5-2 Internal procedures

The function used to load the main GUI interface of the OLIGOSPECIFICITYSYSTEM program is OSS(). The functions Oss(), oss(), OligoSpecificitySystem(), oligospecificitysystem() or Oligospecificitysystem() are shortcuts to the OSS() function (Figure 6). These functions load the general GUI interface that links to the other procedures.

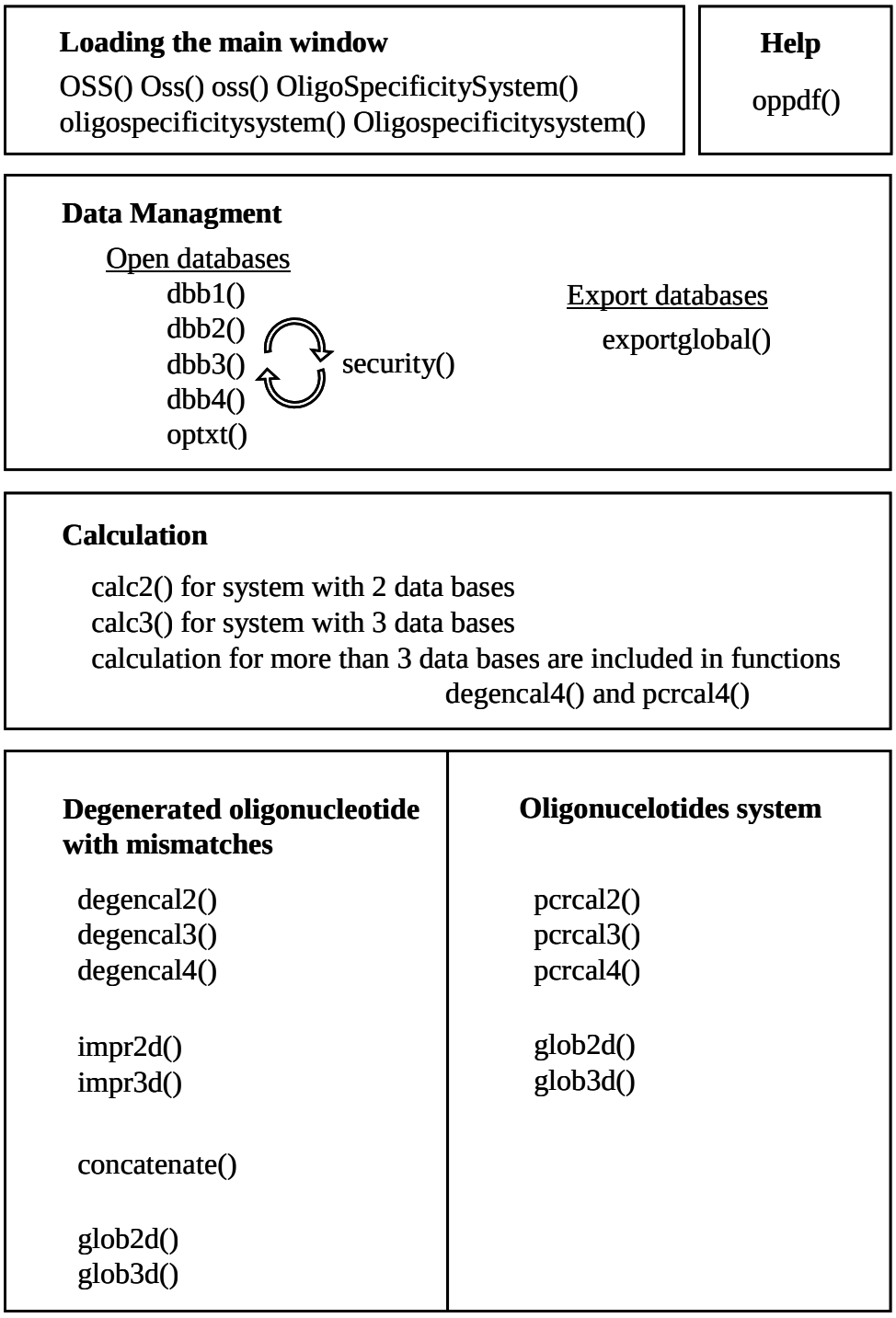


Figure 6. Diagram of functions invoked in OLIGOSPECIFICITYSYSTEM program

Here is a complete list of the functions used in the OLIGOSPECIFICITYSYSTEM program:

- calc2() and calc3() are internal function used to collect information in systems with 2 or 3 databases. calc2() function is used by the functions degencal2(), glob2d(), impr2d() and pcrcal2() whereas calc3() function is used by the functions degencal3(), glob3d(), impr3d() and pcrcal3().
- concatenate() is the function used by the CONCATENATE DATABASES OF DEGENERATE OLIGONUCLEOTIDE procedure of the DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES section.
- dbb1(), dbb2(), dbb3() and dbb4() are functions used to LOAD procedures of the DATA MANAGEMENT section of the 1st, 2nd, 3rd and the following databases respectively.
- degencal2(), degencal3() and degencal4() are functions respectively used by the GLOBAL EFFICIENCY procedure of the DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES section for 2, 3 and more than 3 databases.
- exportglobal() function is used by the EXPORT procedure of the DATA MANAGEMENT section.
- glob2d() and glob3d() functions are used by the ALL INFORMATION ON OLIGONUCLEOTIDE DATABASES procedure in DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES and OLIGONUCLEOTIDES SYSTEM sections for 2 and 3 databases, respectively.
- impr2d() and impr3d() are the functions respectively used by the IMPROVEMENT OF DEGENERACIES procedure of the DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES section for 2 and 3 databases.
- OnOk() is the function used to load the SPECIES SEPARATOR of the DATA MANAGEMENT section.
- oppdf() is used for opening the PDF document of the HELP button.
- optxt() is used by the EXAMPLE DATABASE procedure of the DATA MANAGEMENT section.
- pcrcal2(), pcrcal3() and pcrcal4() are functions respectively used by the GLOBAL EFFICIENCY procedure of the OLIGONUCLEOTIDES SYSTEM section for 2, 3 and more than 3 databases.
- security() is an internal error procedure which verify that databases are loaded. This function is used in OSS function.

5-3 Principles of calculation

5-3-1 Preliminarily calculations

All executive functions of the OLIGOSPECIFICITYSYSTEM program first compute for the i loaded oligonucleotide databases a system of i^2-i+1 equations with i^2-i+1 x_i variables. Each variable x_i corresponds to the sequences set of i^{th} oligonucleotide database (Figure 7). The $|x_i|$ coresponds to the cardinale of the x_i set, e.g. the number of elements (sequences) of the x_i set.

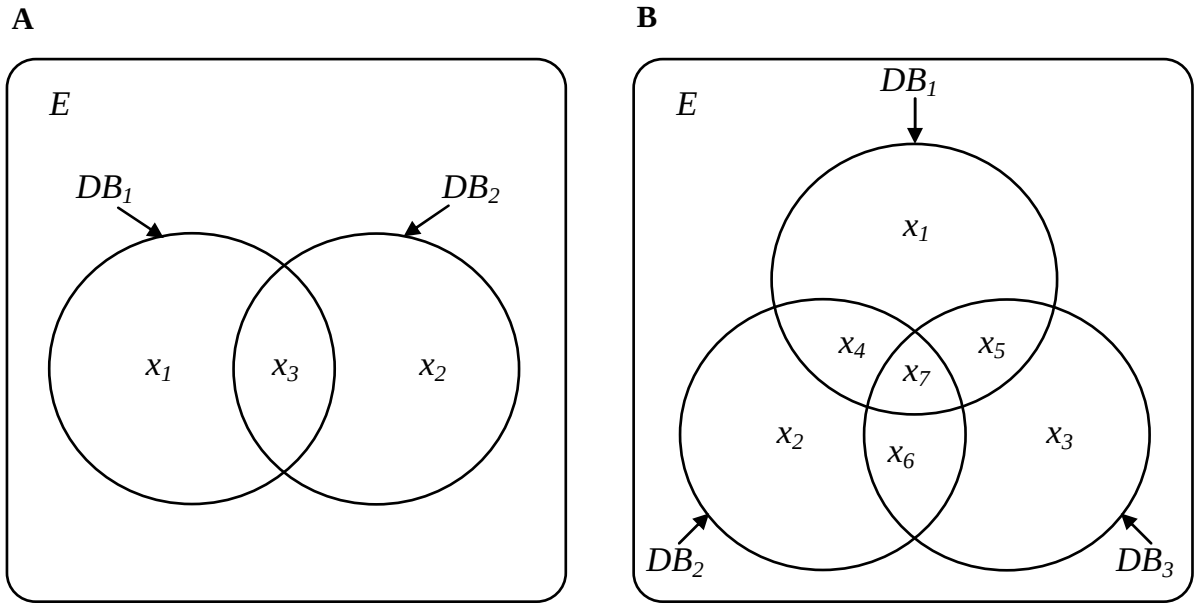


Figure 7. Diagram of the oligonucleotide systems with 2 (A) and (3) databases.

Here are examples of some equation systems with 2 (S_1) and 3 (S_2) oligonucleotide databases (Figure 7).

$$(S_1)\begin{cases} |x_1|+|x_3|=|DB_1| \\ |x_2|+|x_3|=|DB_2| \\ |x_1|+|x_2|+|x_3|=|DB_1 \cup DB_2| \end{cases}$$

$$(S_2)\begin{cases} |x_1|+|x_4|+|x_5|+|x_7|=|DB_1| \\ |x_2|+|x_4|+|x_6|+|x_7|=|DB_2| \\ |x_3|+|x_5|+|x_6|+|x_7|=|DB_3| \\ |x_1|+|x_2|+|x_4|+|x_5|+|x_6|+|x_7|=|DB_1 \cup DB_2| \\ |x_1|+|x_3|+|x_4|+|x_5|+|x_6|+|x_7|=|DB_1 \cup DB_3| \\ |x_2|+|x_3|+|x_4|+|x_5|+|x_6|+|x_7|=|DB_2 \cup DB_3| \\ |x_1|+|x_2|+|x_3|+|x_4|+|x_5|+|x_6|+|x_7|=|DB_1 \cup DB_2 \cup DB_3| \end{cases}$$

The preliminarily calculations consist in solving (S_1) or (S_2) and thus determining $|x_i|$ with i from 1 to 3 in (S_1) or with i from 1 to 7 in (S_2).

5-3-2 Calculation for degenerate oligonuleotide

The calculation of GLOBAL EFFICIENCY procedure in DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES section consisted in determining $|DB_1 \cup DB_2|$ equivalent to $\sum_{i=1}^{3}|x_i|$ in (S_1) and determining $|DB_1 \cup DB_2 \cup DB_3|$ equivalent to $\sum_{i=1}^{7}|x_i|$ in (S_2).

The calculation of IMPROVEMENT OF DEGENERACIES procedure in DEGENERATE OLIGONUCLEOTIDE WITH MISMATCHES section consists in determining $|DB_1 - DB_1 \cap DB_2|$ and $|DB_2 - DB_1 \cap DB_2|$ in (S_1).

It consists in determining $|DB_1 - DB_1 \cap DB_2 - DB_1 \cap DB_3 + DB_1 \cap DB_2 \cap DB_3|$, $|DB_2 - DB_2 \cap DB_1 - DB_2 \cap DB_3 + DB_1 \cap DB_2 \cap DB_3|$ and $|DB_3 - DB_3 \cap DB_1 - DB_3 \cap DB_2 + DB_1 \cap DB_2 \cap DB_3|$ in (S_2).

5-3-3 Calculation for oligonucleotides system

The calculation of GLOBAL EFFICIENCY procedure in OLIGONUCLEOTIDES SYSTEM section consists in determining $\left|DB_1 \cap DB_2\right|$ equivalent to $\left|x_3\right|$ in $\left(S_1\right)$ and $\left|DB_1 \cap DB_2 \cap DB_3\right|$ equivalent to $\left|x_7\right|$ in $\left(S_2\right)$.

Acknowledgement

The authors are grateful to developers of other R packages: tcltk and tkrplot.

References

Ludwig W, Strunk O, Westranm R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lübmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer K-H (2004) ARB: a software environment for sequence data. *Nucleic Acids Research* **32**(4)**:** 1363-1371

R development Core Team (2009) *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing.