

# WORMBASE CONVERTER

Documentation



# Summary

<b>I – Installation</b> .....	4
A – Versions available .....	4
B – Installation .....	4
1 – <i>Under Windows</i> .....	4
2 – <i>Under Linux</i> .....	4
3 – <i>Other operating systems and source code</i> .....	5
<b>II – Configuration</b> .....	6
A – The Configuration Wizard .....	6
B – The WormBase Updater .....	6
C – Update settings .....	7
1 – <i>General parameters</i> .....	7
2 – <i>FTP Server parameters</i> .....	7
3 – <i>Database parameters</i> .....	7
4 – <i>FTP Server parameters for the “letter”</i> .....	7
<b>III – WormBase updates</b> .....	9
A – Initial installation of the releases .....	9
B – Creation of a new package of WormBase releases .....	9
C – Operation updates .....	10
1 – <i>Update process</i> .....	10
2 – <i>Obtaining the list of new releases</i> .....	10
3 – <i>Retrieval of the geneIDs</i> .....	10
4 – <i>Identification of the differences between two releases of WormBase</i> .....	10
5 – <i>Retrieval of all changes in the identifiers of the genes</i> .....	10
6 – <i>Getting the date of a new release</i> .....	11
7 – <i>Identification of the changes in the name of the genes between two releases</i> .....	11
8 – <i>Installation of the new release</i> .....	11
D – Errors, LOG files and how to correct errors .....	12
1 – <i>Identification of errors</i> .....	12
2 – <i>Error description</i> .....	12
3 – <i>Correction of errors</i> .....	13

<b>IV – The WormBase Converter</b> .....	14
A – Graphic interface .....	14
B – General process of a conversion .....	15
1 – <i>Conversion process</i> .....	15
2 – <i>Identification of IDs in the input list</i> .....	15
3 – <i>Identification of the WormBase release in the input list</i> .....	15
4 – <i>Conversion of all genes into WormBase ID</i> .....	16
5 – <i>Conversion / Reverse-conversion</i> .....	16
6 – <i>Conversion of genes into the output ID</i> .....	16
7 – <i>Display of the results</i> .....	17
C – Conversion and Reverse-conversion .....	17
1 – <i>Conversion</i> .....	17
2 – <i>Reverse-conversion</i> .....	17
<b>V – Development</b> .....	18
A - Programming languages & Dependencies .....	18
B – License .....	18
<b>VI – Contact</b> .....	18

# I – Installation

## A – Versions available

The WormBase Converter was developed with Python, using Tkinter for the graphic interface, and Perl. These programming languages allow a good portability between the different operating systems. Thus, this program is available for Microsoft Windows (compiled version), Linux, and can be used and/or compiled under other operating systems with the source code.

Different installation profiles (versions) of the software are available :

- Local version :
  - This installation profile is to use the software on an independent computer.
  - The program runs entirely on the same computer (conversion and updates).
  - The user has a full access to the software.
- Server version :
  - This version must be installed on a server.
  - It centralizes the WormBase releases, and allows the update of all the clients automatically.
- Client version (Administrator) :
  - This version must be installed on a client computer.
  - The program runs on the computer, but updates are downloaded from a server.
  - The user has a full access to the software, and can modify the server settings.
- Client version :
  - This version must be installed on a client computer.
  - The program runs on the computer, but updates are downloaded from a server.
  - The user can view but not modify the server settings.

## B – Installation

### 1 – Under Windows

To install the software under Microsoft Windows, just download the file corresponding to your use (installation profile) for Windows, and extract it on the computer.

Launch the “Configuration wizard” [see II, II.A].

Finally, you can install the **initial** releases of WormBase [see III.A].

NOTE :

In the case of a network architecture (server/client), the initial releases can only be installed on the server. All the clients will be updated automatically.

### 2 – Under Linux

To install the software under Linux, download the “Sources” corresponding to your use (installation profile), and extract it on the computer. To run the program, you must install the dependencies manually [see V.A] or download and install the Debian Package “Dependencies.deb”.

Launch the file “INSTALL.sh” to configure the program and install the initial releases (if necessary). [see II, II.A and III.A].

NOTE :

In the case of a network architecture (server/client), the initial releases can only be installed on the server. All the clients will be updated automatically.

NOTE :

To install the software on a server without graphic interface, launch the file "INSTALL.sh" with arguments:

- 1<sup>st</sup> argument: Path of the package with the WormBase releases
- 2<sup>nd</sup> argument (optional): equals "yes" to install the ACeDB files from the package

### 3 – Other operating systems and source code

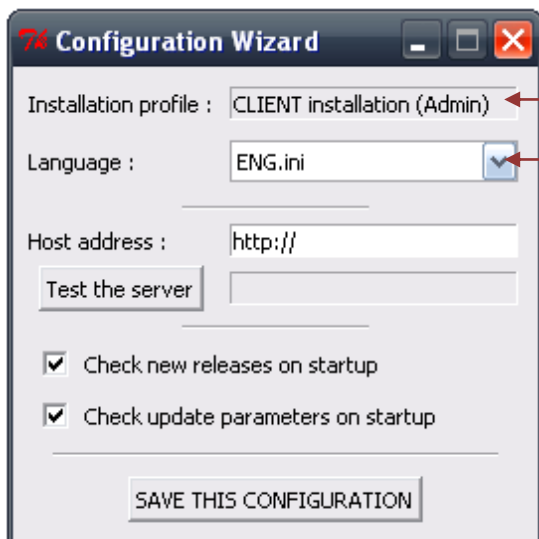
For the other operating systems, you can use the source code with Python and Perl interpreters, or recompile the program under these systems.

However, the WormBase Converter uses and depends on other libraries that must be installed before the execution or the compilation of the program. See the section [V.A] to find the list of these libraries.

## II – Configuration

### A – The Configuration Wizard

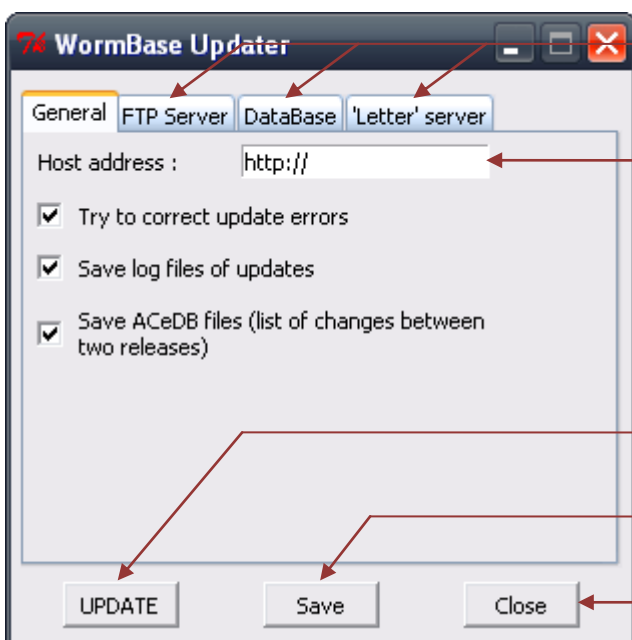
According to the installation profiles, the configuration wizard is more or less complete. For more information, see the section [I.A]. Here an example with the 'client version (administrator)'.



- Installation profile installed
- Language used in the WormBase Converter
- Address of the server (in the case of a network installation). You can test the address entered by clicking on the button.
- Options to check the update settings and/or the new releases available on the server (in the case of a network installation).

### B – The WormBase Updater

According to the installation profiles, the access to the update settings can be limited. For more information, see the section [I.A]. Here is an example with the 'client version (administrator)'.



- Update settings for the FTP server of WormBase, the database and the server of the "letter". [See II.C]
- Address of the server, where the WB Converter is installed [see II.C.1]
- General parameters of updates [see II.C.1]
- Launch an update (if necessary) locally or on the server (in the case of a network architecture)
- Save the changes in the settings locally and on the server (in the case of a network architecture)
- Stop any updates (in the case of a local installation), and close the window

## C – Update settings

### 1 – General parameters

<i>Host address</i>	Only available for a network installation. Address of the host, where the server version is installed.
<i>Try to correct update errors</i>	If an error occurs during the update, a second attempt to retrieve data information is performed.
<i>Save log files of updates</i>	Save all steps of updates (including errors) in two files : <ul style="list-style-type: none"><li>- “wormbase/temp/UPDATES_LOG.txt”</li><li>- “wormbase/temp/UPDATES_LOG_ACeDB.txt”</li></ul> These files can help you to correct manually the errors that occurred during updates.
<i>Save ACeDB files (list of changes between two releases)</i>	Save all “Revision History” information retrieved from the database (ACeDB) during an update. Information is saved in the folder “wormbase/temp/ACeDB”.

### 2 – FTP Server parameters

The “FTP Server” defines the official FTP server of WormBase. It is used to check new WormBase releases, and to retrieve information about these releases (for example the gene IDs).

<i>Host &amp; Port</i>	Address of the FTP server of WormBase, and the port to use to connect to this server. (default : <i>ftp.wormbase.org</i> )
<i>Login &amp; Password</i>	Identifiers used for the connection to the server.
<i>Regular expression for GeneIDs files</i>	Regular expression to define the path of the geneIDs archive, and the format used by the filenames.

If one or more parameters are not necessary, just leave the box empty.

#### INFORMATION : *Regular expression*

String with a specific format, which contains special characters. For example, the points (“.”) must be preceded by an anti-slash (“\.”), and the word “<release>” will be replaced by the number of a WormBase release automatically.

### 3 – Database parameters

The WormBase database (ACeDB : *A C. elegans DataBase*) is used to collect the “Revision History” information about all genes, that is to say the changes that appeared in the name of these genes between the different releases.

<i>Host &amp; Port</i>	Address of the database, and the port to use to connect to it. (default : <i>mining.wormbase.org</i> )
<i>Login &amp; Password</i>	Identifiers used for the connection to the database.

If one or more parameters are not necessary, just leave the box empty.

### 4 – FTP Server parameters for the “letter”

The “letter” is a file available on a specific FTP server that describes a WormBase release and its changes compared to the previous release. This “letter” is used to get the availability date of a release.

<i>Host &amp; Port</i>	Address of the FTP server where to find the “letter”, and the port to use to connect to this server. (default : <i>ftp.sanger.ac.uk</i> )
<i>Login &amp; Password</i>	Identifiers used for the connection to the server.
<i>Regular expression for GeneIDs files</i>	Regular expression to define the path of the “letter”, and the format used by the filenames.

If one or more parameters are not necessary, just leave the box empty.

INFORMATION : *Regular expression*

String with a specific format, which contains special characters. For example, the points (".") must be preceded by an anti-slash ("\."), and the word "<release>" will be replaced by the number of a WormBase release automatically.



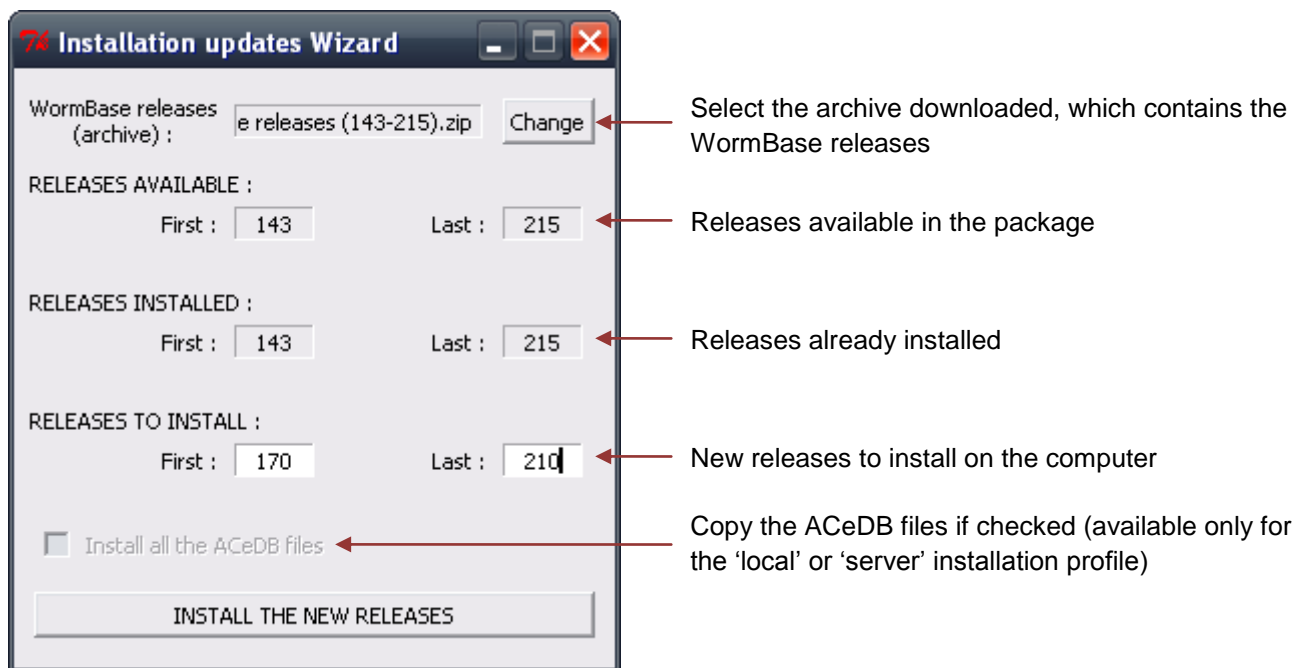
### III – WormBase updates

#### A – Initial installation of the releases

Before the first use of the WormBase Converter, it's necessary to install the WormBase releases.

Download the package of the WormBase releases on the computer.

Launch the “Extract releases”, and follow the instructions :



**NOTE :**

In the case of a network architecture (server/client), the initial releases can only be installed on the server. All the clients will be updated automatically.

#### B – Creation of a new package of WormBase releases

It's possible to create a new “initial WormBase releases” package to share updates and simplify the installation of the WormBase Converter on new computers.

Firstly, create a new directory with three folders named “ACeDB”, “geneEvolvs” and “geneIDs”.

Secondly, copy the files of the desired releases from the “wormbase/geneIDs” folder into the “geneIDs” folder previously created. Do the same with the folders :

- “wormbase/geneEvolvs” => “geneEvolvs”
- “wormbase/temp/ACeDB” => “ACeDB”

Finally, compress the three folders (“ACeDB”, “geneIDs” and “geneEvolvs”) in a ‘tar.gz’ file without password.

**NOTE :**

The package must contain all releases from the first to the last one desired. It's impossible to skip a release.

The ACeDB files are optional, but an error will occur if somebody wants to install them from this package.

## C – Operation updates

### 1 – Update process

Update of the WormBase releases is performed automatically in several steps :

- Obtaining the list of the releases available [see III.C.2]
- Retrieval of the geneIDs of the new release [see III.C.3]
- Identification of the differences between the new release and the previous [see III.C.4]
- Retrieval of all changes in the name of genes in the new release [see III.C.5]
- Getting the date of the new release [see III.C.6]
- Identification of the changes in the name of genes between the new release and the previous [III.C.7]
- Copy of the new files, and installation of the new release [see III.C.8]

#### NOTE :

In the case of a network architecture (server/client), the new releases available on the server are automatically downloaded on the client computers.

### 2 – Obtaining the list of new releases

To find the new WormBase releases, the program checks all releases available on the FTP server, using the regular expression defined in the parameters [see II.C.2]. A simple comparison is then performed between releases available and releases installed.

#### NOTE :

Only the WormBase releases higher or equal to WS143 are checked on the FTP server.

### 3 – Retrieval of the geneIDs

GeneIDs of a new release are downloaded from the FTP server of WormBase, using the regular expression defined in the parameters [see II.A.2].

Initially, the geneIDs are compressed in a Gzip file (\*.gz). An extraction of this archive is performed to get the file with the different geneIDs of the release :

- WormBase ID [e.g. WBGene00000020] : the identifier of a unique gene, used by WormBase
- Gene Name [e.g. cnc-2] : the name given to a gene
- Gene Sequence Name [e.g. R13A5.9] : the name given to a gene with a specific format

### 4 – Identification of the differences between two releases of WormBase

Firstly, a simple comparison of the WormBase IDs in the new release and the previous is performed to determine the genes added or removed.

A file called “diff\_WSxxx-WSyyy” (xxx : number of the previous release, yyy : number of the new release) is generated, with the following format :

	WormBase ID	[tab]	Modification	[r\n]
e.g.	WBGene00000020		ADD	
e.g.	WBGene00000021		DEL	

#### INFORMATION : *Modification & Special characters*

The modification “ADD” means that the gene was added in the new release, and “DEL” that the gene was deleted in the new release.

[tab] is a tab (not spaces) - [r\n] are the characters ‘carriage return’ and ‘newline’

### 5 – Retrieval of all changes in the identifier of the genes

Secondly, on the basis of the file of differences between two releases, a script written in Perl queries the WormBase database (ACeDB) to get all information about the changes in the name of genes between these two releases. The address of the database and information for the connection can be modified in the parameters [see II.C.3].

The WormBase database uses the AQL language and the querying can be performed with Aceperl, only available with Perl. Three data are retrieved from the database :

- the changes in the name of a gene :  

```
request:select g->Identity[History][Version_change][5] from g in
class Gene where g = "<WormBase ID>";
```
- the dates of the changes :  

```
request:select g->Identity[History][Version_change][2] from g in
class Gene where g = "<WormBase ID>";
```
- more details about the changes, if necessary :  

```
request:select g->Identity[History][<change>] from g in class Gene
where g = "<WormBase ID>";
```

Several changes (*terms* used in the database) in the name of genes are considered :

- *Created & Imported* : the gene appears in the new release
- *Resurrected* : the gene, removed in a previous release, reappears in the new release
- *Killed & Made in transposon* : the gene is removed in the new release
- *Split from* : the new gene is created by the splitting of a existing gene
- *Split to* : the gene is divided into two or more new genes
- *Merged into* : the gene fuses with a existing gene in the new release

Several errors can occur during the retrieval of these data. To know how to correct the different errors, see the section [III.D].

## 6 – Getting the date of a new release

The date of the new release is necessary to determine what changes, in the list of changes retrieved from the database, occur exactly between the new release and the previous. This date is written in a “letter” file on a FTP server (see [III.C.4] to modify the parameters), which is downloaded during the update.

## 7 – Identification of the changes in the name of the genes between two releases

A relationship between the date of the new release, the date of the previous release, and the date of changes in the name of genes allows identifying which changes occurred between the two releases.

### ALGORITHM :

The dates of the new and the previous release are not enough to determine which changes appeared between the two releases. So, several methods are used to correct this problem (see the source code for more information).

## 8 – Installation of the new release

To use the new release in the WormBase Converter, files created during the update are copied in several folders of the program :

- the geneIDs are saved in the folder “wormbase/geneIDs”
- the file with all changes between the two releases is saved in the folder “wormbase/geneEvolvs”

Depending of the options selected, some other files are saved :

- the LOG files are saved in the folder “wormbase/temp”
- the ACeDB file is saved in the folder “wormbase/temp/ACeDB”

All other files created during the update are removed from the folder “wormbase/temp”.

Finally, the numbers of the new releases are added to the “config.ini” file.

## D – Errors, LOG files and how to correct errors

### 1 – Identification of errors

To identify and correct errors occurred during the WormBase updates, it's necessary to activate the following options [see II.A.1] :

- save log files of updates
- save ACeDB files (list of changes between two releases)

The LOG files are saved in the folder “wormbase/temp”, and the ACeDB file in the folder “wormbase/temp/ACeDB”.

All information about updates (progression, errors, warnings ...) is written in these files, and allows identifying when and/or where an error occurred during the process.

#### ADVICE :

After each update, you should check these two files to see if errors occurred during the update, and correct these errors.

### 2 – Errors description

Three types of error exist, and are saved in the LOG files :

- errors due to the program
- errors due to the FTP servers
- errors due to the WormBase database

#### Errors due to the program :

<i>Unable to open the geneIDs file from release WSxxx</i>	The program cannot open the file that contains all the geneIDs of a release. This error can occur if the program does not read access to the file.
<i>The number of genes changed between WSxxx and WSyyy is more than 10 000</i>	This is a warning. It means than the update will be very long because there are many changes in the name of genes between the two releases.
<i>Unable to install the new release</i>	The program cannot install the new release on the computer. This error can occur if the program does not have write access to the folder “wormbase/geneEvol” and/or “wormbase/geneIDs”.
<i>Update interrupted by the user</i>	The user has stopped the update, for example by clicking on the “Close” button or by closing the program (for a local installation).
<i>Cannot open file (wormbase/temp/diff_WSxxx-WSyyy)</i>	The program cannot find the file with the differences between the gene lists of the two releases. This error can occur if the program does not have read access to the folder “wormbase/temp”.
<i>Cannot create file (wormbase/temp/ACeDB_WSxxx-WSyyy)</i>	The program cannot create the file with all changes in the names of genes. This error can occur if the program does not have write access to the folder “wormbase/temp”.

#### Errors due to the FTP servers :

- |  |   |
|--|---|
| <i>Unable to connect to the server</i> | The program cannot connect to the FTP server.<br>This problem can occur if the server is temporary unavailable, or if the |
|--|---|

	FTP server settings are incorrect.
<i>Unable to check the new releases</i>	The program cannot obtain the list of the WormBase releases available. This problem can occur if the server is temporary unavailable, or if the FTP server settings are incorrect (especially the regular expression).
<i>Unable to download the geneIDs file from release WSxxx</i>	The program cannot download the archive that contains the geneIDs of the WormBase release. This problem can occur if the server is temporary unavailable, or if the FTP server settings are incorrect (especially the regular expression).
<i>Unable to connect to the server</i> OR <i>Unable to download the "letter" file from release WSxxx</i>	The program cannot download the "letter" file to get the date of the new release. This problem can occur if the server is temporary unavailable, or if the FTP server settings are incorrect.

Errors due to the database :

<i>It's impossible to retrieve the changes of several genes.</i>	Some errors occurred during the retrieval of the changes in the name of genes. These errors have been sent by the database.
<i>Querying the database</i> [...] .... error : <desc>	An error has been sent by the database server. The description of the error is written instead of <desc>.
<i>The number of dates is different than the number of changes.</i>	In the database, the number of dates of changes is different from the number of changes. This error can only be corrected manually.

### 3 – Correction of errors

In the case of an error due to the program or to the FTP servers, check the settings and the rights to the folders and files of the program. And finally, restart the update.

**NOTE :**

If the same FTP error always occurs, it is likely that there is a problem with the FTP server. For example, a file required for the update is not present on the server. Only a correction on the server can resolve the problem.

In the case of an error due to the database, it's necessary to correct it manually. See the specific tutorial available at "help/Tutorial - Correction updates.pdf".

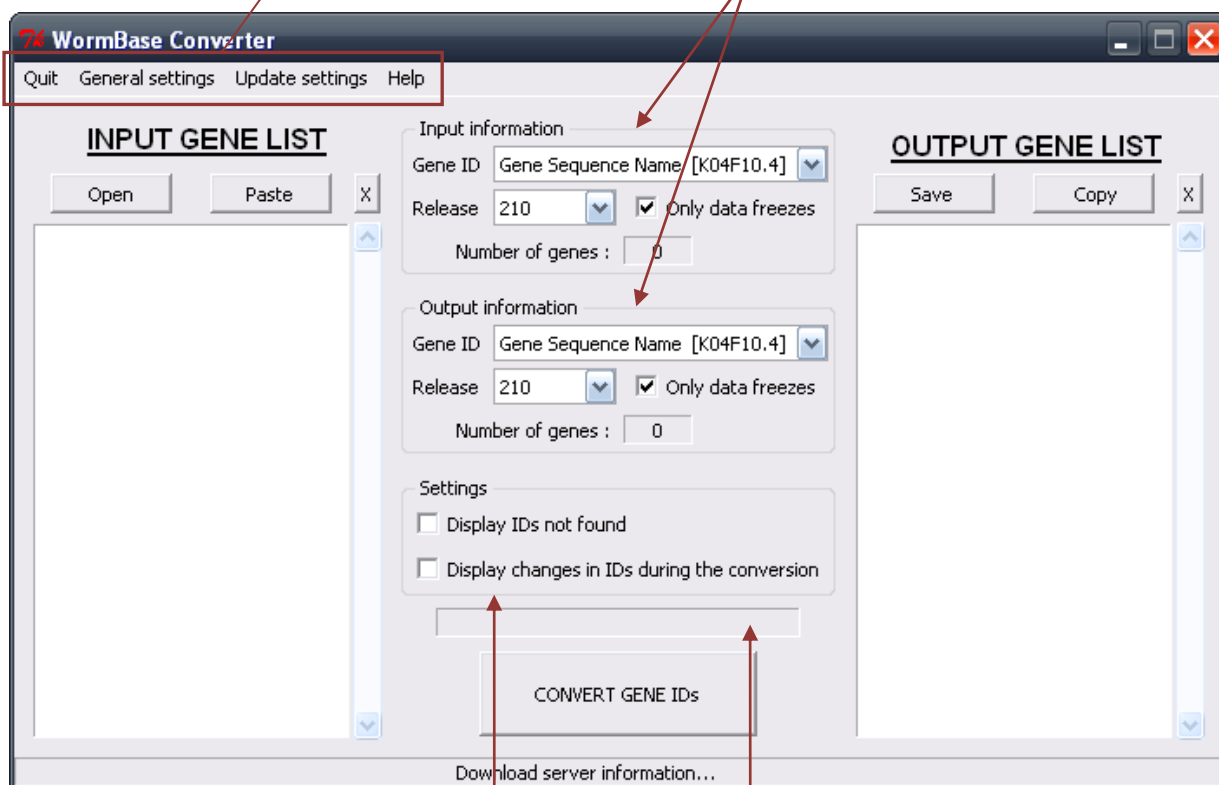
## IV – The WormBase Converter

### A – Graphic interface

**MENU :** Settings of the software, of updates and help. See below for details.

**INPUT INFORMATION :** Gene ID and release used in the input list

**OUTPUT INFORMATION :** GeneID and release wanted after conversion



**INPUT GENE LIST :** Paste or open a file which contains the genes that you want to convert.

**SETTINGS :** Options to obtain more details about the conversion. See below.

**PROGRESS BAR :** Progression (steps) of the conversion.

**OUTPUT GENE LIST :** The list of genes after conversion (or reverse-conversion). You can copy or save this list.

**Menu :**

- **Quit :** Exit the program and any updates (in the case of a local installation)
- **General settings :**
  - **Language :** Language used in the WormBase Converter interface
  - **Default settings :** Check (or not) the options of the conversion by default
  - **Installation profile :** Change the installation profile between “Client” and “Local”
- **Update settings :**
  - **Launch the WormBase Updater :** Open the window to change the updates settings [see II.B] or to update the installation (locally or on the server)
  - **Update settings – Check new releases on startup :** Check if there are new WormBase releases available on the server (in the case of a client installation)
  - **Update settings – Check new parameters on startup :** Check if the update parameters have been changed on the server (in the case of a client installation)
- **Help :**
  - **Tutorial :** Tutorial to use the WormBase Converter quickly
  - **Help :** Open this documentation
  - **About :** Information about the program

Settings (more details about the conversion):

- *Display IDs not found* : Display a window, after conversion, to show the list of genes not found in the input release selected, and/or genes which have not the ID wanted in the output release.
- *Display changes in IDs during the conversion* : Display a window, after conversion, to show all changes in the name of genes between input release and output release.

## B – General process of a conversion

### 1 – Conversion process

The conversion (or reverse-conversion) can be separated in several steps:

- Identification of IDs used in the input list [see IV.B.2]
- Identification of the release used in the input list [see IV.B.3]
- Conversion of all genes of the input list into WormBase ID [see IV.B.4]
- Conversion or reverse-conversion of genes ID from input release to output release [see IV.B.5]
- Conversion of genes of the output list into the ID wanted [see IV.B.6]
- Display the result of the (reverse-)conversion and optional details windows [IV.B.7]

### 2 – Identification of IDs in the input list

Different IDs can be used in the input list :

- *WB Gene ID* : The WormBase Gene ID used by the WormBase database
- *Gene Sequence Name* : The name given to a gene with a specific format
- *CDS Sequence Name* : The name given to a CDS (coding sequence)
- *Transcript Sequence Name* : The name given to a transcript
- *Gene Name* : The name given to a gene

NOTE :

If one of these IDs is selected, only genes of the input list corresponding to this ID will be converted. The other genes will be ignored.

However, it's possible to convert a gene list which contains more than one ID by selecting '- *Unknown / Mix* -' as GeneID. All the genes in the input list will be converted (provided that they exist).

Moreover, if you are not sure of the ID used in the input list, select '- *Unknown / Mix* -' to detect automatically the IDs of the list.

ALGORITHM :

The rules to detect automatically the IDs used in the input list are :

- the ID begins by "WB" => WormBase ID
- the ID has two points => Transcript Sequence Name
- the ID has one point and the last character is alphabetic => CDS Sequence Name
- the ID has one point, no dash, and the last character is not alphabetic => Gene Sequence Name
- all other IDs => Gene Name

### 3 – Identification of the WormBase release in the input list

It's possible to select a specific WormBase release ("freeze" or not) for the input gene list. Releases available correspond to the releases installed on the computer. To filter the list to see only releases "freeze", check the box '*Only data freezes*'.

However, if you don't know the WormBase release used in the input gene list, select '*Unknown*' to detect this release automatically. In the case where there is only one release possible, this one is selected automatically for the conversion ; but if there are several possible releases, a window will open containing the five most probable releases and you will have to choose the one you want to use.

**ALGORITHM :**

Automatic detection of the release used in the input list is based on the percentage of genes of the input list found in the WormBase releases. The following method is applied :

- search for the two best “frozen” releases (with the highest percentages)
- calculation of the percentage in all releases between these two best “frozen” releases
- sorting the results in descending order according to percentage
- sorting the results in descending order according to release number
- the five “best” releases will be displayed in a window (see above)

**NOTE :**

In the case of a mix of releases in the input list, only one can be used for the conversion. So, all genes not found in the input release selected will not be converted.

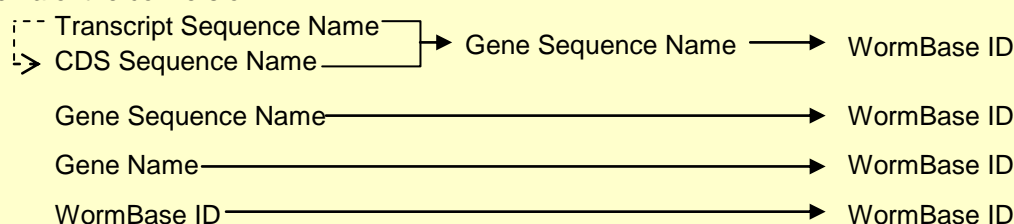
**4 – Conversion of all genes into WormBase ID**

An automatic conversion of all gene IDs into WormBase ID is performed before the real conversion, using the input release selected.

All genes not found in the input release will not be converted, and will be displayed in the window “*IDs not found*” if option is checked.

**ALGORITHM :**

Schema of the conversion :

**5 – Conversion / Reverse-conversion**

If the input release is smaller than the output release, a conversion is performed [see IV.C.1].

If the input release is higher than the output release, a reverse-conversion is performed [see IV.C.2].

All the changes in the name of genes, during the (reverse-)conversion, are saved in the file “temp/conversion.txt”, and can be displayed in a window if the option “*Display changes in IDs during the conversion*” is checked.

**NOTE :**

If you want to keep the list of all changes, check the option mentioned above, and click on “Save”. The file “temp/conversion.txt” is erased at each new (reverse-)conversion.

The file with all changes (“conversion.txt”) uses the following format:

	Conversion	WB ID before conversion	Conversion	WB ID after conversion
e.g.	170 --> 171	WBGene00011457	<i>Merged_into</i>	WBGene00011456
e.g.	171 --> 170	WBGene00011456	<i>Merged_into (reverse)</i>	WBGene00011457

**6 – Conversion of genes into the output ID**

Finally, a conversion of the gene IDs is performed to give the output ID wanted (“WormBase ID”, “Gene Sequence Name” or “Gene Name”), using the output release selected.

The genes which do not have the requested ID will not appeared in the output list, but can be displayed in the window “*IDs not found*” if the option is checked.



## 7 – Display the results

After (reverse-)conversion, the converted genes appear in the output list without duplicates, and the genes not converted can be displayed in the optional window “*IDs not found*”. All changes in the name of genes can also be displayed in the optional window “*Changes in IDs during the conversion*”.

### NOTE :

There is no correspondence between lines in the input list and the output list. If you want to follow changes for a specific gene, use the option “*Changes in IDs during the conversion*”.

## C – Conversion and Reverse-conversion

### 1 – Conversion

A “conversion”, between each release is performed when the input release is smaller than the output release. It uses the information from the WormBase database (files in folder “wormbase/geneEvol”).

The changes analyzed are :

- *Merged/Split to* : the gene ID is replaced by another one
- *Split from* : a new gene ID appears
- *Killed (and Made in transposon)* : the gene ID is removed
- *No change* : the gene ID stays the same

Moreover, a specific system is used to manage the genes annotated as *Resurrected*.

### 2 – Reverse-conversion

A “reverse-conversion” between each release is performed when the input release is higher than the output release. It uses the information from the WormBase database (files in folder “wormbase/geneEvol”), and defines the inverse of this information.

The changes taken into account are :

- *Merged (reverse)* : the gene ID is replaced by two gene IDs (the same gene and the source of merging)
- *Split to (reverse)* : the gene ID is replaced by the source of splitting
- *Split from (reverse)* : the gene ID is replaced by the source of splitting
- *Created (reverse) (and Resurrected (reverse))* : the gene ID is removed
- *No change* : the gene ID stays the same

A specific system allows also managing the genes *Killed* as genes *Resurrected* (= *Killed (reverse)*).

## V – Development

### A - Programming languages & Dependencies

The WormBase Converter uses different libraries that must be installed on the computer running the program.

Under Windows, the compiled version of the program already integrates these libraries.

Under Linux, the file “Dependencies.deb” allows a simple installation of these libraries.

In other cases, you must install manually these libraries (see the list below).

Python modules :

<u>Description</u>	<u>Module name</u>
Python	<i>python (2.5 to 2.7)</i>
Tkinter	<i>tk8.5</i>
Tcl	<i>tcl8.5</i> <i>tcllib</i>
Tkinter for python	<i>python-tk</i>
Table list object	<i>tklib</i>

Perl modules :

<u>Description</u>	<u>Module name</u>
	<i>perl (5.10 or higher)</i>
Perl	<i>libcache-cache-perl</i> <i>liberror-perl</i>
AcePerl	<i>libace-perl</i>

### B – License

The WormBase Converter is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

The WormBase Converter is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License (<http://www.gnu.org/licenses/>) for more details.

## VI – Contact

WormBase Converter, conceived by Jonathan Ewbank, was entirely written by Aurélien Griffon, under the supervision of Ilka Engelmann.

CIML – Centre d’Immunologie de Marseille-Luminy

Parc Scientifique & Technologique de Luminy

Case 906

13288 MARSEILLE cedex 09

FRANCE

Contact : <[ewbank@ciml.univ-mrs.fr](mailto:ewbank@ciml.univ-mrs.fr)>