

Allele Name Translation Tool (ANTT) v 0.5.0 User Guide

1. [Overview](#)
2. [Requirements](#)
3. [Using the ANTT](#)
4. [Limitations and Troubleshooting](#)
5. [Release Notes](#)

1. Overview

The Allele Name Translation Tool (ANTT) is intended to translate HLA allele names between various naming-conventions and nomenclature rules. In particular, it has been designed to translate between alleles written under May 2002 to March 2010 naming conventions (e.g., A*01010101) and alleles written under the April 2010 naming conventions (e.g., A*01:01:01:01).

The ANTT package includes four applications -- *setup.exe*, *ANTT.exe*, *ANTT_GUI.exe*, and *ANTT_FileManager.exe*.

The *setup.exe* application will install the .NET environment on Windows XP and Vista systems. To use the ANTT on MacOSX, BSD, or Linux systems, install the Mono open source .NET development framework (see section 2, Requirements).

ANTT.exe is a console/command line version of the Allele Name Translation Tool that can be run from the command line (or by double-clicking in Windows systems). The console application allows the ANTT to be incorporated into a larger data-manipulation pipeline or for datasets to be translated in batches.

ANTT_GUI.exe provides a graphical user interface (GUI) for the Allele Name Translation Tool. The GUI application makes it easy to navigate through your file system to select configuration files and input data files in different directories.

ANTT_FileManager.exe allows the user to view, edit, and save configuration files. While configuration files are text files, errors may easily be introduced by attempting to edit them manually. The ANTT File Manager makes it easy to make changes to configuration files while minimizing opportunities to introduce errors.

The File Manager also allows the user to generate individual translation files from a single translation source file, and to repair input data files that have column number discrepancies between rows.

The ANTT package also includes *config.ini*, a basic configuration file, *example_dataset.txt*, a sample ANTT input-data file, and the *translation_files* directory, which contains the translation files used by the ANTT to translate between allele naming conventions at each locus. The translation files distributed in this directory are derived from the translation source file available at http://hla.alleles.org/data/txt/Nomenclature_2009.txt. [Alternative translation tables](#), and associated configuration files, are also distributed with the ANTT.

2. Requirements

To use the ANTT, you must have an input-data file, which contains columns of allele data, a configuration (*.ini) file, which tells the ANTT what to look for in the input file, and a set of translation files, which tell the ANTT how

to translate each allele.

The ANTT runs in the [Microsoft .NET framework](#) (version 3.5 SP1) in Windows XP and Vista, should run natively in Windows 7, and can be run on Apple OSX, BSD, and Linux platforms using the [Mono open source .NET development framework](#) (version 2.4). Specific information about installing these frameworks can be found in the ANTT_README.txt file distributed with the ANTT.

2.1 File Definitions

The ANTT input files are tab-delimited text files that contain columns of data (data organized in rows are not supported). Each column can include allele name data for only one locus, and each column of data must have a header (the field-name) identifying the contents of that column. These field-names must be included in the configuration file. Both allele-data and non-allele data can be included in the input file. Ambiguous allele strings (e.g., A*01010101/0102/0103) can be included in the input file, but the character that separates the alleles in a string must be defined in the configuration file. A missing data character can also be defined in the configuration file. Allele data that include characters (e.g., single or double quotes) other than those that constitute allele-names or missing data or that identify separate alleles in strings will not be recognized.

The configuration file can be set up using the ANTT File Manager. This program lets you import and edit settings from another configuration file, or enter all of the settings yourself. You must specify the full name of the new configuration file, the path to the translation files (this can be done by identifying one of the files in the correct folder, or by entering the name of the folder it is in the same directory as the ANTT), the possible fields in the input file (all fields do not need to be in one file, but this allows you to use the same configuration file for multiple input files), and the translation file that corresponds to a particular locus. In addition, you can specify the character or characters used to separate ambiguous alleles in a string (e.g., A*01010101/0102/0103 represents three alleles separated by slashes, while A*01010101,0102,0103 represents the same three alleles separated by commas), the character or characters used to represent missing data, if your input file contains ambiguous alleles, or uses a particular missing data symbol.

The ANTT translation tables are tab-delimited text files that contain at least two columns. The left-most column contains the reference allele names, which should correspond to the alleles in the input file. The right column should contain the translated allele names, which will appear in the output file. Each column must have a header row containing column names. Locus prefixes should be included in the names of the alleles in both columns. Translation files for translating from non-colon delimited allele names to colon-delimited allele names are included in the translation_files folder. These files are derived from the list available at http://hla.alleles.org/data/txt/Nomenclature_2009.txt.

Translation files can be generated from a translation source file using the ANTT File Manager. The translation source file is a tab- or space-delimited text file that contains the allele-translation correspondences for all alleles and loci of interest in two columns (e.g., http://hla.alleles.org/data/txt/Nomenclature_2009.txt).

3. Using the ANTT

The ANTT can be run from the command line using the ANTT.exe application, or through a graphical user interface (GUI) using the ANTT_GUI.exe application.

3.1 Graphical User Interface

To use the ANTT's GUI in Windows, double-click the ANTT_GUI.exe file. To use the ANTT's GUI under Mono,

enter "mono ANTT_GUI.exe" from the command line. To start translating, select "Translate Alleles" from the "Actions" pull-down menu. You will be prompted to identify the configuration file to use for the translation in a separate window, followed by a prompt to identify the input file (also in a separate window). Translation progress and error messages (e.g., allele names that could not be translated, or truncated allele names that were translated) will be displayed in the "Translation Log and Information" window.

To translate a dataset that contains alleles written under a newer nomenclature to an older nomenclature using the GUI, select "Back (Reverse) Translate" from the "Settings" pulldown menu and then select "Translate Alleles" from the "Actions" pull-down. To return to standard translation, uncheck this option under "Settings".

3.2 Log Settings

To include all instances of allele issues (e.g., truncations, or untranslatable alleles) select "Verbose Logging" in the "Settings" pull-down menu. The default behavior is to report only the first instance of each allele issue.

To report the cell position (row and column) of each reported allele issue, select "Cell Reporting" in the "Settings" menu.

To report the sample identifier for a given allele issue, select the "Sample ID Reporting" option in the "Settings" menu. Sample IDs in the left-most 'id' column will be reported.

Unselect each settings option to return to the corresponding default ANTT behavior.

3.3 Command Line

To use the ANTT from the command line in Windows, enter "ANTT" followed by up to four parameters separated by spaces (ANTT parameter0 parameter1 parameter2 parameter3). To use the ANTT from the command line under Mono, enter "mono ANTT.exe parameter0 parameter1 parameter2 parameter3".

Parameter0 can include the following six options:

"-h" will generate this help message.

"-a" will generate an "About the ANTT" message.

The ANTT will quit after the -h or -a options are used.

"-b" will reverse translate from new to old nomenclature.

"-v" activates 'Verbose Logging' of all allele issues, otherwise only one instance of each issue is logged.

"-c" reports the Cell (row and column) of each logged issue.

"-i" reports the sample identifiers (IDs) of each logged allele issue.

IDs are derived from the first (left-most) 'id' column.

-b, -v, -c, and -i can be combined (e.g., -bvc, -vi) as well.

Parameter1 must be the name of the configuration (.ini) file.

Parameter2 must be the name of the input-data file.

Parameter3 must provide a name for the file into which translated data will be written (the translated-data file).

All parameters are optional when the ANTT is started from the command line. Parameter1 and parameter2 can be entered and parameter3 omitted, or parameter1 can be entered alone, and parameter2 and parameter3 omitted, or all three parameters can be omitted. If nothing is entered, the user will be prompted for parameter0

or parameter1.

If parameter1 is not provided, you will be prompted for the name of a valid config.ini file. If parameter2 is omitted, you will be prompted for the name of a valid input-data file. If parameter3 is omitted, the name of the translated-data file will be generated using the name of the input-data file, with "-translation" appended to the input-data file name (-translation.txt). Translation progress and error messages (e.g., allele names that could not be translated, or truncated allele names that were translated) will be displayed in the console window.

Allele translation will begin after valid configuration and data-input filenames have been entered. When the translation is completed, the ANTT generates two output files. Translated data are saved with the name [input file name]-translation.txt. These translation.txt files can be used as inputs for the ANTT. The ANTT also generates a log file named [input file name]-translation_log.txt, which includes all translation progress and error messages.

Alternative Translation Tables

If your data consists of primarily full-length allele names (e.g., A*01010101), then most translations will proceed very rapidly using the translation tables in /translation_tables. However, if your data contains many truncated allele names (e.g., A*01, A*0101 or A*010101), the ANTT will take longer to complete the translation as it checks the validity of each truncated allele name. If you are certain about the validity of the truncated allele names in your dataset, you can translate your data using the translation tables in the /complete_truncation_translation_tables directory and the associated config_complete_translation.ini file. Using these translation tables, the translation of truncated allele names will not appear in the translation log file, and the data will be translated faster.

Similarly, if your data contains allele names that end in a lower-case g, signifying alleles that encode the same antigen recognition sequence (ARS) as described by Cano et al., Hum Immunol. 2007 68(5):392-417, and others, you can expand these groups to slash-delimited strings of full-length v2.* allele-names using the translation tables in the /ARS_translation_tables directory and the associated config_ARS.ini file. These strings can then be translated to their v3.2.0 cognates using the translation tables in /translation_tables.

4. Limitations and Troubleshooting

While the ANTT has been designed with flexibility in mind, so as to accommodate many allele-data storage formats, there are limits on the data storage formats that can be translated using the ANTT. For example, **allele-data stored in rows rather than columns cannot be translated using the ANTT.**

4.1 Truncated Alleles

Alleles truncated as described below cannot be translated unless modified translation tables are used.

When using the translation tables that are distributed with the ANTT, allele names that have been truncated in certain ways will not be translated. Alleles that have been truncated to omit a leading '0' (e.g. 1010101 instead of 01010101) will not be translated, because 1010101 is not a recognized allele-name.

Allele names that incorporate '00' or 'XX' to represent allelic ambiguity will not be translated (e.g., DQB1*0500 or DQB1*05XX). In these cases, the allele name should be truncated to two digits (e.g., DQB1*05).

These types of truncated allele names will be passed to the translated data output file unchanged, and their

appearance in the dataset will be noted in the translation log file.

However, it is possible to translate these types of truncations, using a modified translation file that explicitly identifies the truncated allele and the allele to which it should be translated.

4.2 *Variation in the characters used to separate ambiguous alleles*

The configuration file can identify only one character (or string of characters) as being used to separate the alleles in an ambiguous allele string. Therefore only ambiguous allele strings using that character can be parsed for a given configuration file. If you include ambiguous allele strings that use different characters to separate the alleles in a single input-data file, only one type of string will be parsed by the ANTT.

4.3 *Unexpected carriage return/line feed characters*

Some spreadsheet programs using non-western character sets may generate tab-delimited text files that include characters that are recognized by the ANTT as carriage return (CR) or line feed (LF) characters before the end of a line. This results in the ANTT sending the user an error message that the different numbers of columns are present in the rows in the input data file. You can try and repair these files using the Repair Input Data File function of the ANTT File Manager. In general, the ANTT expects the lines in data-input files to end with a CR followed by a LF.

4.4 *Problems While Running the ANTT*

The ANTT checks to make sure that the configuration file and input-data files specified by the user exist, and that they are valid filetypes. The most likely cause of an inability to translate an input-data file is an error specifying the path to the translation tables, or the name of a translation table, in the configuration file. The ANTT_FileManager.exe program can be used to minimize errors in generating new configuration files.

5. *Release Notes*

ANTT v0.5.0

Bug Fixes:

-- Translation files that include only 1 allele-name pair are now handled correctly. Thanks to Tom Smith for the report.

-- Column is not translated if the associated translation file is empty or includes only a header. Thanks to Tom Smith for the report.

-- Allele names with odd-numbers of digits that appear to be potentially truncated (e.g. 01011) are no longer translated.

-- Both the ANTT and ANTT_GUI now record "Performing 'back-translation' from newer to older nomenclatures." in the log file, when back-translation is performed.

-- Both the ANTT and ANTT_GUI now record the name of the data file being translated in the log file.

-- ANTT no longer prompts users to "Press return to end ..." twice after requesting 'help' or 'about' information.

New Features:

-- Output log options have been added to the ANTT and ANTT_GUI.

-- The following Log Settings can be specified via checkboxes in the 'Settings' pulldown menu added to the ANTT_GUI:

--- 'Verbose Logging': reports all instances of allele issues in the log file; allele issues are not reported in the ANTT console, or the ANTT_GUI translation log and information window. Under the default (Terse Logging) mode, only the first instance of an issue for any allele is reported. Note, Verbose Logging mode can significantly

increase the runtime of the tool if your data contains many truncated allele names.

--- 'Sample ID Reporting': The value in the left-most column with the 'id' header will be reported for each logged allele issue. Thanks to the Australian Red Cross Blood Service for this suggestion.

--- 'Cell Reporting': The row and column of each allele issue will be reported for each logged issue.

-- In the ANTT, these settings are activated by passing -v (verbose), -i (ID) or -c (cell) parameters, or combinations thereof (e.g., -iv) in the parameter0 position.

-- The default reporting mode is to not report sample IDs or cells for logged issues, and to only log each issue once.

-- Forward or back translation can be set in the Translation Settings section of the 'Settings' menu in the ANTT_GUI. Forward translation is the default mode.

Other Issues:

-- An alertbox that appeared when one-character long allele-names were included in input-data has been removed.

-- The Back-Translate option has been removed from the 'Actions' menu of the ANTT_GUI. See New Features above for more information.

-- The translation log entry indicating that the translation file contains multiple mappings for certain alleles, "Translation file <filename> contains multiple entries for the same allele:....", has been changed from an "Error" message to a "Warning" message. Thanks to Tom Smith for the report.

Translation Tables:

-- The translation tables in /translation_tables have been updated using the September 20, 2010 version of the Nomenclature_2009.txt file from ftp.ebi.ac.uk.

-- New translation tables for the expansion of version 2.* nomenclature allele names encoded as g-groups (e.g., A*0101g) of alleles that encode a common antigen recognition sequence (ARS) to corresponding version-2 allele strings (e.g., A*01010101/01010102N/0104N/0122N/0132/0134N/0137) are included in the /ARS_translation_tables directory. An associated configuration file (config_ARS.ini) is distributed as well. Thanks to Pierre-Antoine Gourraud for developing these.

-- New translation tables including complete left-truncations (e.g., A*01010101 is truncated to A*01, A*0101 and A*010101, and A*01:01:01:01 is truncated to A*01, A*01:01 and A*01:01:01) of all final version 2.* nomenclature allele names and their corresponding v3.2.0 allele names are included in the /complete_truncation_translation_tables directory. These tables also include translations of deleted/changed v1.* and v2.* allele names to their v3.2.0 cognates. An associated configuration file (config_complete_truncation.ini) is distributed as well.

ANTT File Manager v0.5.1

Bug Fixes:

-- When input data files less than 1KB in size are repaired, a *-repaired.txt file is now generated.